Understanding Glycolysis in *Escherichia coli*: a Systems Approach using Nuclear Magnetic Resonance Spectroscopy

Johann Josef Eicher

Dissertation presented for the degree of Doctor of Philosophy
in the Faculty of Science at Stellenbosch University

Supervisor Prof. Johann M. Rohwer
Co-supervisor Prof. Jacky L. Snoep

December, 2013

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 22 November 2013

Summary

This dissertation explores the behaviour and regulation of central carbon metabolism in *Escherichia coli* K12 W3110 under fermentative microaerobic conditions. To achieve this, an integrative systems modelling approach was adopted, which is introduced in Chapter 1 along with a review of metabolism in *E. coli*. An open-source software suite NMRPy, developed using the Python programming language, is presented in Chapter 2. NMRPy provides a host functions for basic processing, analysis and visualisation of Nuclear Magnetic Resonance (NMR) spectroscopy data. In addition to this, NMRPy offers specialised functions for the deconvolution of arrayed reaction time series, which proved indispensable to the research presented in this dissertation. NMRPy presents an easy to use, extensible tool for both routine and advanced use. In Chapter 3, a novel methodology is presented which was developed for the effective and comprehensive determination of enzyme kinetic parameters for systems biology using NMR. In contrast to traditional enzyme kinetic assay methods, this new methodology is less labour-intensive and yields significantly more information per experiment. By fitting kinetic equations to real time NMR data, dynamic changes in substrates, products and allosteric modifiers are quantified and allowed to inform the parameter fitting procedure. These data contain information on cooperative substrate binding, reversibility, product inhibition and allosteric effects. The proposed methodology is applied to the study of the first two enzymes of the glycolytic pathway. In Chapter 4, the construction, parameterisation and validation of a number of kinetic models of glycolysis in *E. coli* under microaerobic conditions is detailed. To model the lower half of glycolysis, a similar technique was adopted as in Chapter 3, in which models representing the reactions from triosephosphate isomerase to pyruvate kinase were parameterised by fitting them to a collection of ^{31}P NMR reaction time series. This approach extends the methodology to enzyme sub-networks, yielding data that encompass the full complexity of the network regulatory interactions. The verified kinetic models were subjected to scrutiny, the results of which are presented in Chapter 5. The value of the modelling approach is demonstrated by the ease with which cumbersome *in vivo* experiments can be performed *in silico*. A structural analysis of the model topology was conducted, elucidating the elementary flux modes of fermentative glycolysis in *E. coli*, and identifying a futile cycle around PEP carboxylase and PEP carboxykinase. Model steady-state behaviour and control properties were explored *in silico* under various degrees of ATP demand and oxygen availability and a number of hypotheses are presented, explaining the regulation of free energy in *E. coli*, and the metabolic responses of *E. coli* to changing redox demands. Amongst other things, the results demonstrated that the glucose importing phosphoenolpyruvate:phosphotransferase pathway controlled glycolytic flux, and that under microaerobic conditions *E. coli* is able to regulate redox balance not only by balancing flux between acetate and ethanol, but also by altering the balance of flux between acetate and lactate at the pyruvate formate lyase/lactate dehydrogenase branch point. This study demonstrates the value of an integrated computational and experimental systems approach to exploring biological phenomena.

Opsomming

In hierdie proefskrif word die gedrag en regulering van die sentrale koolstofmetabolisme in *Escherichia coli* K12 W3110 onder fermenterende mikro-aërobiese toestande ondersoek. Dit is moontlik gemaak deur 'n geïntegreerde stelsel-modelleringsbenadering, wat in Hoofstuk 1 bekendgestel word. Díe hoofstuk verskaf ook 'n oorsig van die metabolisme in *E. coli*. 'n Oopbron-kodepakket NMRPy, wat in die programmeringstaal Python ontwikkel is, word in Hoofstuk 2 beskryf. NMRPy verskaf 'n aantal funksies vir die basiese verwerking, analise en visualisering van Kern-Magnetiese Resonansie (KMR) spektroskopiese data, sowel as gespesialiseerde funksies vir die dekonvolusie van opeenvolgende reaksie-tydreekse. Hierdie funksionaliteit was onontbeerlik vir die verdere navorsing in hierdie proefskrif. Hoofstuk 3 beskryf die ontwikkeling van 'n nuwe metodiek vir die omvangryke bepaling van ensiem-kinetiese parameters vir sisteembiologie, deur van KMR gebruik te maak. In teenstelling tot tradisionele ensiem-kinetiese essai-metodes, is hierdie nuwe metodologie minder arbeidsintensief en lewer dit beduidend meer inligting per eksperiment. Deur die kinetiese vergelykings op tydsafhanklike KMR data te pas, word dinamiese veranderinge in substrate, produkte en allosteriese effektors gekwantifiseer en hierdie inligting gebruik in die passingsprosedure. Die data bevat inligting oor koöperatiewe substraatbinding, omkeerbaarheid, produkinhibisie en allosteriese effekte. Die voorgestelde metodologie word toegepas op die karakterisering van die eerste twee glikolitiese ensieme. In Hoofstuk 4 word die konstruksie, parameterisering en validering van 'n aantal kinetiese modelle van glikolise in *E. coli* onder mikro-aërobiese toestande uiteengesit. Die waarde van die modelleringsbenadering lê in die gemak waarmee omslagtige *in vivo* eksperimente *in silico* uitgevoer kan word. Om die onderste helfte van die glikolitiese pad te modelleer word 'n soortgelyke tegniek as in Hoofstuk 3 gebruik. Modelle van die reaksies vanaf triosefosfaat-isomerase tot by pirovaat-kinase is geparameteriseer deur dit op 'n versameling ^{31}P KMR-tydreekse te pas. Hierdie benadering brei bostaande metodologie uit tot ensiem-subnetwerke en genereer data wat die volle kompleksiteit van regulerende interaksies in die netwerk insluit. Die geverifieerde modelle word in Hoofstuk 5 noukeurig ondersoek. 'n Strukturele analise van die modeltopologie word onderneem om die elementêre fluksie-modes van fermentatiewe glikolise in *E. coli* te verklaar, sowel as om 'n futiele siklus rondom fosfoënolpirovaat karboksilase en fosfoënolpirovaat karboksikinase te identifiseer. Die bestendige-toestandsgedrag en kontrole-eienskappe word *in silico* ondersoek onder toestande van verskeie ATP beladings en suurstofbeskikbaarheid. 'n Aantal hipoteses word voorgelê, wat die regulering van vry energie in *E. coli*, sowel as die metaboliese reaksies van *E. coli* onder veranderende redoks-vereistes kan verklaar. Onder andere dui die resultate daarop dat die fosfoënolpirovaat:fosfotransferase sisteem (wat verantwoordelik is vir glukose-opname in die sel) die glikolitiese fluksie beheer en dat *E. coli* onder mikro-aërobiese toestande die redoksbalans nie net tussen asetaat en etanol kan reguleer nie, maar ook die deur wysiging van die fluksie-balans tussen asetaat en laktaat rondom die pirovaat-formiaat-liase/laktaatdehidrogenase vertakkingspunt. Hierdie studie toon die waarde van 'n geïntegreerde rekenaarmatige en eksperimentele sisteembenadering om biologiese verskynsels te ondersoek.

Acknowledgements

A PhD thesis is no mean feat, and without the unwavering hopefulness and care of friends, family and loved ones this thesis would not have been completed. It is as much their work as mine.

Firstly, I would like to extend heartfelt thanks to my supervisor Prof. Johann Rohwer for his constant encouragement along the way, especially in the times when I found myself without a compass. Watching him think and work has taught me once again to pay attention to detail. Additionally I am grateful to my co-supervisor Prof. Jacky Snoep for constructive conversations and key suggestions to this research. Without Prof. Jannie Hofmeyr taking the time to get to know me as a floundering Masters graduate, and sharing his wisdom and inspiration with me, I would not have taken up this task.

I would also like to thank my family for their patience, care and support, and Jana for lovingly and unquestioningly redirecting me to the source of our hope as many times as was necessary.

I thank all the members of the Department of Biochemistry at Stellenbosch University, and especially those belonging to the Triple-J research group for their unhesitating friendship and acceptance of an alien. Very helpful conversations and fun were shared with C-J Sidego, Theo van Staden, Dr. Chris Visser, Carl Christensen, Lana Keet, Waldo Adams, Dr. Danie Palm, Francois Du Toit and many others. Dr. Jaco Brand and Elsa Malherbe were indispensable to this work.

I am grateful to the National Research Foundation for financial support during this research, and for their willingness to accommodate unexpected obstacles. Special thanks must be extended to Sinazo Peter for painstakingly fulfilling all of the administrative requirements.

Aure entuluva.

Abbreviations regularly used in this text

α-MG	α-methyl glucoside
2PG	2-phosphoglycerate
3PG	3-phosphoglycerate
ACE	phosphate acetyltransferase + acetate kinase
ADH	alcohol dehydrogenase
ADP	adenosine diphosphate
ALD	fructose 1,6-bisphosphate aldolase
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BPG	1,3-bisphosphoglycerate
DHAP	dihydroxyacetone phosphate
DTT	dithiothreitol
EI	PTS enzyme I
EIIA	PTS enzyme IIA
EIICB	PTS enzyme IICB
ENO	enolase
ETOH	ethanol
F6P	fructose 6-phosphate
FBP	fructose 1,6-bisphosphate
FID	free induction decay
FRD	fumarate reductase
FUM	fumarase
G3PDH	α-glycerol 3-phosphate dehydrogenase
G6P	glucose 6-phosphate
G6PDH	glucose 6-phosphate dehydrogenase
GAP	glyceraldehyde 3-phosphate
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GLC_{ex}	external glucose
GLK	glucokinase
HPr	PTS enzyme heat-stable protein
LCMS	liquid chromatography/mass spectrometry
LDH	l-lactate dehydrogenase
MCA	Metabolic Control Analysis
MDH	malate dehydrogenase
MQO	menaquinone
MQOH	menaquinol
NAD^+	β-nicotinamide adenine dinucleotide (oxidised)
NADH	β-nicotinamide adenine dinucleotide (reduced)
NMR	Nuclear Magnetic Resonance spectroscopy
PDHC	pyruvate dehydrogenase complex
PEP	phosphoenolpyruvate
PFK	phosphofructokinase
PFL	pyruvate-formate lyase
PGD	6-phosphogluconate dehydrogenase
PGI	phosphoglucose isomerase
PGK	phosphoglycerate kinase
PGL	6-phosphogluconolactonase
PGM	phosphoglycerate mutase
PIPES	piperazine-N,N-bis[2-ethanesulfonic acid]
PK	pyruvate kinase
PPC	phosphoenolpyruvate carboxylase
PTS	phosphoenolpyruvate-phosphotransferase system
PYR	pyruvate
RF	radio frequency
SDA	Supply-Demand Analysis
SUCC	succinate
TIM	triosephosphate isomerase

Contents

1 Systems biology and the modelling of microbial physiology

Summary

This chapter presents a review of the background literature relevant to the current study. Topics covered include a discussion of the value of approaching the study of biological phenomena with a systems mindset, and a review of the paradigms, strategies and tools adopted by systems biologists for exploring metabolic networks, with a focus on the study of central carbon metabolism in *Escherichia coli*. A number of questions are posed at the end of the chapter which served to guide the research presented in later chapters.

1.1 Introduction

Escherichia coli is the *de facto* model organism by virtue of the significant amount of scientific attention it has received over time, and has been described as a laboratory "workhorse" [1], being key to recombinant molecular biological laboratory techniques and a favoured protein expression host (90 % of structures deposited in the Protein Data Bank used *E. coli* as an expression host [1]). Indeed *E. coli* has provided scientists with some landmark discoveries and insights, including amongst others: bacterial conjugation [2], bacteriophages and phage genetics [3, 4], the mechanisms of evolution studied in Richard Lenski's famous long-term experiment [5]; and after the sequencing of the *E. coli* genome in 1997 [6], a long series of constraint-based whole genome modelling attempts have been performed on *E. coli* [7]. Much of this work, however, has been reductionistic and has failed to account for the nuanced characteristics of biological networks. In contrast therefore, the approach adopted in this study to further illuminate the central carbon metabolic pathway of this much studied organism is integrative, constructive and with a systems-level scope.

The advances in the systems approach to studying biological phenomena have been motivated by the realisation that reductionism—the "divide and conquer" scientific method which has predominated since the Enlightenment [8], an effective strategy for cataloguing and characterising elementary parts—is unable to account for the complex emergent properties of systems [9]. A core technique of systems biology is mathematical modelling, which allows hypotheses to be tested rapidly *in silico* which may not be amenable to *in vitro* or *in vivo* experimentation due to cost or the lack of experimental means. The field of systems biology involves a paradigm shift: elements are viewed in the context of the system, and their individual contributions to the system's behaviour are quantified [10]. Viewing elements thus in the context of the system has been significantly aided by the development of frameworks such as Control Analysis by Kacser and Burns [11] and Heinrich and Rapoport [12].

It is within this paradigm that we seek to understand the behaviour and regulation of central carbon metabolism in *E. coli*, particularly in the context of fermentative growth at low O_2 availability producing the mixed acid fermentation products. The kinetic modelling approach adopted in this dissertation provides the platform to understand the complexities of the glycolytic network of *E. coli* under fermentative microaerobic conditions on three levels: topologically, the stoichiometric relationships between the reactions of glycolysis reveal fundamental flux distribution achievable by the system (e.g. by using Elementary Mode Analysis [13]); in terms of control properties, the regulatory contribution of each reaction and metabolite or block of reactions is quantifiable (e.g. through the use of Metabolic Control Analysis [14], and Supply-Demand Analysis [15]); and particular to kinetic models, the dynamic behaviour over time is observable.

Additionally, two tools relevant to this study and to systems biology in general, were developed and are presented in this dissertation. The first tool, NMRPy, is a software suite for the rapid and convenient processing of Nuclear Magnetic Resonance spectroscopy (NMR) data. NMRPy is a necessary and useful tool, which provides a degree of extensibility and adaptability not offered

by the current *milieu* of NMR processing packages, and is particularly configured to handle NMR reaction time series. The second tool presented is a methodology for approximating enzyme kinetics using NMR reaction time series. The methodology exhibits a number of advantages over traditional enzyme kinetics techniques, allowing more data to be extracted from each experiment, including information relevant to product-inhibition, reversibility of reactions, cooperative substrate binding, and allosterism. It is our hope that these two offerings will be welcome experimental contributions to the research community, especially in light of the recent revival of interest in enzyme kinetics [16], and the advances in NMR technology which suggest that it will become increasingly useful to the field of enzyme kinetics and metabolomics [17–19].

A review of the techniques and biological background relevant to this dissertation is presented below.

1.2 Modelling flux

Two polarised approaches to modelling biological systems have emerged in the literature: an inductive "top-down" approach, in which elementary interactions are inferred from general system properties; and a deductive "bottom-up" approach, which aims to predict complex systemic behaviour from a basis of mechanistically-detailed constitutive elements [20]. The extrema of these two strands of biology have been caricatured as follows [21]:

1. the Data-Mining Approach: In this popular approach, which is often used in molecular biology and the medical sciences, the presence or correlation of biological phenomena is tested by identifying qualitative differences and correlations in datasets (e.g. gene-knockouts, wild-type vs. mutant, microarray data). Generally this type of modelling has access to large datasets and is strictly descriptive, not necessarily containing mechanistic explanations for biological phenomena.

2. the Systems Approach: The systems approach on the other hand is mechanistic and dynamic, describing the change in state of biological systems usually described by models comprised of differential equations. Models are both mechanistically descriptive and predictive, allowing the elucidation of complex phenomena.

The "bottom-up" approach adopted by this study is constructive and by definition modular, allowing the integration of various sub-models into a larger systemic model (e.g. the whole-cell modelling of *Mycoplasma genitalium* [22]). This strategy has been employed in modelling various systems, including, amongst others, yeast glycolysis [23, 24], sucrose accumulation in sugarcane culm [25, 26], erythrocyte glycolysis [27], 2,3-bisphosphoglycerate metabolism in the erythrocyte [28], *Plasmodium falciparum* glycolysis [29], the thioredoxin system in *E. coli* [30], and *Trypanosome* glycolysis [31, 32].

The growth of "omics"-technology has enabled researchers to produce vast amount of biological data which, in the absence of adequate interpretation and analysis, is in danger of redundancy.

Thorough scientific utilisation of these data must involve the elucidation of underlying mechanisms. Merely topological descriptions of biological data in the absence of an attempt at explaining the emergence of patterns of dynamic system behaviours from basis elements is a shortcoming of similar magnitude to creating a road map whilst neglecting to understand traffic patterns [33].

Not only is it essential for the myriad of data from all the strata of current biology to be understood within their respective domains (e.g. genetics, transcriptomics, proteomics, metabolomics), but the relationships across domains need to be understood to achieve a holistic vision of the functioning of cellular life, and to grow out of a "high-data era" into a "high-information era" [34]. A partial implementation of the integrated approach is the linkage of stoichiometric data with genetic sequence information [35]. This "phenomic" strategy, essentially a form of functional genomics, defines and predicts genotype-phenotype relationships but alone it remains merely a structural description of the system being studied as it makes scant comment on dynamic system properties such as metabolite flux and its control (much like the static "road map" idea above), and is thus lacking in predictive power. Against this background, harmonisation of the various scientific domains has emerged as a defining prerogative of systems biology which attempts to exploit a systems understanding of biological networks by revealing their key properties, including network topologies and the structural relationships between constituents, dynamical time variation in systems, and control methods responsible for the regulation of dynamic behaviour [36].

One of the great milestones for future biology will be to model the entire cell *in silico*. This would ultimately involve the integration of the genome (and epigenome) sequence, gene-products and a map of the metabolic network in a dynamic simulation capable of responding to changes in environmental and genetic conditions [37]. A project with this goal was started in 1996 and is known as the "E-cell project" [38] using primarily *Mycoplasma genitalium* DNA supplemented with *E. coli* genes. A similar project called the "Silicon Cell initiative" [39] adopts a mechanistic approach to modelling by piecing together experimental data and making detailed sub-cellular models available in an online database allowing easy access, evaluation and curation (http://www.jjj.bio.vu.nl) with the goal of '[combining] models of parts of cellular systems and ultimately to construct detailed kinetic models at the cellular level'[39].

Seeking to explain complex emergent biological behaviour, the systems approach involves abstraction of formal systems (mathematical models) from natural systems (real biological networks) which express system properties that can be validated with biological observations of nature [40]. The effort to cure cancer is a salient example of the need to elucidate an immensely complex signalling network which is unlikely to be achieved by a traditional molecular biology approach without understanding the complex behaviour of cancer cells; an ideal application of the *in silico* modelling approach [41].

The goal of flux-based approaches to microbial metabolic modelling is to functionally describe organisms using a reconstructed network of genomic annotations in combination with experimentally determined biochemical and physiological properties of the components of a cell [33].

Genome annotation or gene ontologies aim to identify all of the open reading frames in a specific genetic sequence and assign functions to them. Procedurally, an organism's sequence and functional data are arranged into a "network" by fitting them to a related metabolic map (e.g. The Kyoto Encyclopedia of Genes and Genomes, "KEGG" [42]). This genomic information, which is essentially a dictionary of the elements of a system and a hypothesized set of connections (i.e. a topology), needs to be evaluated experimentally to be able to validate the assignment of reactions and connections to genes as suggested from genome annotation, and to discover reactions that are not salient in the genomic data [33]. The combined genomic and experimental information yields a reconstructed organism-specific network consisting of a list of elements and their connections, as is found in the EcoCyc database [43]. Such "ontologies" provide the scaffold for flux- modelling.

Two approaches to modelling are discussed below. The first category, "top-down" modelling, produces solutions to modelling problems by constraining the solution space using global properties of the system. The second category, "bottom-up" modelling, seeks to assemble kinetic models with equations describing individual reactions, producing a system with emergent properties resembling experimentally observed global behaviour.

1.2.1 "Top-down" methods: modelling within constraints

"Top-down" approaches to modelling flux predominate in the literature though they are by no means exclusive (and are sometimes combined with "bottom-up" strategies) [44–46]. Three approaches deserve discussion:

Metabolic flux analysis and flux-balance analysis

The strategy of assembling kinetic models in a piecewise manner can be hampered by the lack of experimental information available for all the individual enzymes in a given network. The constraint-based approach, rather than calculating the precise behaviour of the constituents of a network using a comprehensive set of independently determined kinetic parameters, reduces the number of achievable metabolic phenotypes that the system can display by imposing selected constraints (uptake and formation rates, mass-balance equations and thermodynamics) on the model and optimising metabolic flux to satisfy a selected objective criterion or to match experimentally-determined flux patterns [45, 47]. This strategy is realised in the technique called metabolic flux analysis (MFA), of which flux-balance analysis (FBA) is a subset. Constraint-based methods such as MFA and FBA begin by representing a system with mass-balance equations which describe the net concentration of a metabolite as the difference between the input and consumption rates at steady state, that is, when $\frac{d[S]}{dt} = 0$; the change of any specific metabolite pool, S, over time is equal to zero and thus production and consumption rates are equal.

MFA is often used in conjunction with FBA [45]. Whereas the goal of FBA is to determine optimal cellular performance, MFA describes the system flux in greater detail by employing experimental data as constraints to estimate internal reaction fluxes, often through the use of

isotope-enriched substrates [45, 48–50]. This strategy does not require kinetic or thermodynamic information as these constraints are included implicitly in the modelled growth rate as part of the energy and biomass requirements [51].

MFA relies on a stoichiometric model accounting for the net change of all metabolites (substrates, products, intracellular metabolites and biomass constituents) in a system in which all reactions considered can be conveniently grouped into matrix notation [45]. The uptake and formation rates as well as isotopic distribution amongst products can be determined experimentally, and fluxes through pathways can be predicted. Mass-balance equations are used to account for the net accumulation of a specific substrate in the bioreactor environment within which these experiments take place, and yield-coefficients are calculated to provide a "macroscopic" assessment of flux by relating the mass of a specific metabolite produced to the mass of substrate consumed. This is similarly assumed to be at a pseudo steady-state, that is, that the fluxes around a specified metabolite (leading to and from a metabolite pool) are equal. As the dilution rate–due to consumption of a metabolite to produce biomass–is much lower than the metabolic consumption rates in experimental time frames, it can be ignored with the result that all fluxes in linear sequence are equal leaving only branch points to be considered [52].

FBA (summarised by Orth *et al.* [53]) operates within thermodynamic constraints specific to the system, and a bounded solution space is defined within which the solution to the set of balance equations lies [48]. It is possible to narrow the solution space within this boundary by the addition of experimental data. The optimised state of the metabolic network is identified using linear optimisation of the solution space; given a set of upper and lower bounds, $\mathbf{Nv} = \mathbf{0}$ is solved for a particular objective function (where $\mathbf{N}$ is the stoichiometry matrix and $\mathbf{v}$ is a flux vector) [53, 54].

FBA thus requires the choice of an optimality criterion by which a realistic steady state of the system can be derived from the possible states in the solution space. Common objective functions optimise biomass production with respect to growth or the production of a specific metabolite (such as ethanol in yeast); minimisation of metabolic adjustment in response to perturbation (such as a gene deletion); or more generally, maximisation of flux through a specific pathway [55].

The result of FBA is a set of fluxes optimising the objective function and as such it does not produce a parameterised model *per se*. However, FBA is a powerful analytical tool that can be used to understand networks that are far too big for "bottom-up" modelling (e.g. genome scale models) [53].

FBA does suffer from a number of shortcomings, namely: (1) the assumption that the optimal solution is a description of reality (i.e. that the organism has evolved optimally with respect to a chosen objective criterion) while cellular and kinetic evolution is a "work in progress" taking place under temporal and physiological constraints [56]; (2) the potential existence of multiple optimal solutions and (3) the existence of unknown constraints such as regulatory mechanisms [48]. The resulting solutions can be incorporated into computational hypotheses to be generated and tested

experimentally under different environmental conditions (e.g. phenotype phase plane analysis [57]) or different genetic conditions (e.g. *in silico* gene deletion studies [58]). It deserves to be mentioned that through a recent development called "CoPE-FBA", the combinatorial explosion of possible solutions encountered in FBA can be diminished and core flux sub-networks can be identified, greatly aiding the interpretation of constraint-based modelling results [59].

FBA is capable of generating flux distributions that match experimental data; however FBA describes a metabolic network only at a specified steady-state. Predicting dynamic changes effectively, requires knowledge of the kinetic properties of the enzyme components and FBA often sacrifices the kinetic details of a network for a global description of input and output data.

An example of constraint-based flux modelling is described in Section 1.6.2: The Chassagnole *et al.* model.

Metabolic pathway analysis

This approach attempts to treat the structure of the metabolic network mathematically and elucidate indispensable (or elementary) metabolic pathways and how they can be manipulated [48]. Metabolic Pathway Analysis (MPA) likewise begins by deriving a stoichiometric model which defines the connectivity of metabolites representing each metabolite pool with an Ordinary Differential Equation: $\frac{d[S_i]}{dt} = \sum_j N_{i,j} v_j$, where S_i is a specific metabolite, $N_{i,j}$ is the stoichiometry coefficient of metabolite S_i produced by reaction j, and v_j are the fluxes that lead to or from the metabolite [60]. A pseudo-steady state in which generating- and consuming fluxes of a metabolite are equal, is once more justified by the fact that growth rates are much slower than reaction rates in metabolic reactions, and the above set of equations can be simplified by representing them as a set of linear equations in matrix form: $\mathbf{N} \cdot \mathbf{v} = \mathbf{0}$ where $\mathbf{N}$ is an $i \times j$ stoichiometry matrix of i metabolites and j reactions, and $\mathbf{v}$ is a flux vector.

The construction of this stoichiometry matrix forms a useful link between mathematical manipulations and biological knowledge. A solution space is defined for all the possible flux distributions that the metabolic network can assume [61]. The solution space (or the null-space at steady state, that is the set of vectors $\mathbf{v}$ for which $\mathbf{Nv} = \mathbf{0}$) can be described by a set of basis vectors (a linearly-independent vector for each dimension of the solution space) which each represent an operative metabolic pathway. As an orthonormal basis of the solution space, the complete set of possible solutions can be constructed from these vectors (*viz.* the entire set of possible network states can be reconstructed from a combination of scaled basis vectors) [35]. These basis vectors are however not necessarily unique and as such may not describe invariant properties of the network, that is, fundamental pathways.

A unique set of vectors can be derived by an alternative treatment which represents reversible reactions as a combination of irreversible forward and reverse reactions [35]. This strategy results in a solution space that is by definition positive, and the intersection of the null-space with this positive orthant has a convex polyhedral cone shape. The edges of this "flux cone" represent

generating vectors, that is, vectors on the edges of the allowable region of flux vectors. Using a linear combination of these generating vectors, any flux vector within the cone can be produced. These generating vectors, though not linearly independent (but "systemically independent"), are exceptionally useful as they represent elementary biochemical pathways which are structurally indispensable for the given network and its specific outcomes; each a different irreducible set of enzymes, none is a subset of the other [35].

Distinction must be made between elementary modes and extreme pathways. An extreme pathway is a basis mode that lies on the boundary of the allowable flux distribution and defines the shape of the polyhedral cone, that is a generating vector from which steady state solutions (or vectors) within the cone or along its faces can be generated [62]. Therefore, the extreme pathways are analogous to a basis set (or coordinate system) in linear algebra. These extreme pathways are unique when reversible fluxes are decomposed into bidirectional fluxes. In a biochemical context elementary modes are defined as "the minimal set of enzymes that could operate at steady state with all irreversible reactions proceeding in the appropriate directions" [62]. However, when the directional constraint is relaxed and reactions are allowed to be bidirectional the result is usually a set of elementary modes which exceeds the dimensions of the null-space. Thus, some of these elementary modes will be extreme pathways (on the vertices of the cone) and some will be within the cone and thus positive linear combinations of extreme pathways [62].

Much like FBA (which is simply the imposition of constraints on the solution space reducing the convex cone to a polyhedral slice [35]), a comparison of elementary pathways or modes will yield important information with regard to the optimal modes of a network and several computational tools have been developed in this regard (e.g. MetaTool [60]). The application of MPA can be seen in the kinetic modelling of futile cycling during sucrose accumulation in sugar cane culm tissue [25]. Elementary flux modes of the sucrose synthesis and accumulation cycle were generated for a kinetic model and used to identify possible states of the system which exhibit futile-cycling, that is, 5 out of 14 elementary modes resulted in synthesised sucrose being degraded in a futile manner. Combining this stoichiometric analysis with kinetic modelling of the network allowed the authors to make quantitative suggestions with the goal of increasing sucrose yield. MPA has similarly been used to optimise flux to particular metabolites in *E. coli* [63], and to explore the metabolism and pathway redundancies of *Haemophilus influenzae* [64, 65].

MPA was employed to elucidate the elementary modes of *E. coli* glycolysis in this study (Section 5.2).

1.2.2 "Bottom-up" methods: constructing kinetic models

Contrary to the "top-down" methods described above, kinetic modelling approaches are more detailed and incorporate experimentally-determined kinetic information in a "reverse engineering" strategy. Unlike constraint-based models, kinetic models are not limited to a particular steady state. This allows for a greater breadth of *in silico* experimentation. The approach taken in

this study is to construct a model with experimentally-determined *in vitro* kinetic parameters for each enzyme. Component enzymes are treated as linking relationships between different pools of metabolites in the system and can be described explicitly by rate equations that can take into account all of the factors contributing to the rate of a particular reaction:

- substrate/product/modifier concentrations
- forward/reverse rates
- equilibrium constants
- cooperative binding
- allosteric inhibition/activation
- substrate/product inhibition

Kinetic information for enzymes has to be sourced from available literature or elucidated experimentally with the ultimate goal of aggregating these mathematical relationships in an experimentally-verifiable model that responds dynamically to environmental perturbations and can be used to make predictions about the given metabolic network. Often *in vitro* kinetic values for enzymes are used that are not necessarily representative of the intracellular condition, but methods are available to generate *in situ* [66] and *in vivo* data [18, 28, 66, 67].

The bacterial phosphotransferase system

To present the process of kinetic modelling, the analysis of the bacterial phosphotransferase system is described [68]. The phosphoenolpyruvate:glycose phosphotransferase system (PTS) is a bacterial system of group-transfer enzymes which sequentially pass a phosphate group from phosphoenolpyruvate (PEP) to a carbohydrate molecule being imported into the cell (reviewed in [69, 70]). The PTS also plays a regulatory role by exhibiting some control over catabolite repression and chemotaxis in response to fluctuating substrate availability [71, 72]. Metabolic networks are typically idealised in biochemistry as a set of individual reactions taking place in well-stirred volumes (thus excluding diffusion effects), catalysed by enzymes which are present at much lower concentrations than their respective substrates and products; additionally, these enzymes are only able to communicate with each other through variable metabolite pools (thus excluding the direct interaction of proteins) [73]. As a group-transfer pathway, the PTS components do not strictly behave like enzymes as the product of any individual phosphotransfer reaction does not yield the protein catalyst in its original state. Rather, the variously phosphorylated components spend significant amounts of time in complexes with each other [68], and being localised in and adjacent the plasma membrane, the PTS is an example of "non-ideal metabolism", and is subject to the channelling effects of membrane-associated enzymes [68, 74, 75].

Several kinetic characterisations of the PTS components were accessible at the time of the study being described, including K_m values of some enzymes, equilibrium constants, association/dissociation constants and overall forward/reverse rate constants. To adequately model the atypical group-transfer activity of the components of the PTS, the phosphotransfer reactions were deconstructed into association and dissociation reactions, and the collected kinetic parameter values were expressed as elementary forward and reverse rate constants. Simple mass-action rate equations were constructed using the elementary rate constants which could be derived by simultaneously solving these equations for each of the five phosphotransfer steps of the PTS. A number of other parameters were necessary, including enzyme concentrations and boundary metabolite concentrations. These parameters were subsequently altered to simulate three experimental environments: *in vivo*; *in vitro* with low protein concentration; *in vitro* with high protein concentration (to mimic effects of macromolecular crowding).

As the model was constructed entirely from experimental *in vitro* kinetic values, it could be verified by comparison with published experimental observations of *in vitro* and *in vivo* fluxes and response coefficients towards changes in key parameters (e.g. protein concentration). As no kinetic parameters are fitted in this approach, the model produced will approximate reality only to the degree that the experimentally-determined kinetic parameters are accurate and representative of the *in vivo* environment.

The model could be seen to mimic the experimental data by comparison with literature values for *in vitro* and *in vivo*:

- fluxes
- flux-changes induced by changes in individual enzyme concentrations
- flux-changes in response to changes in group enzyme concentrations
- macromolecular crowding (induced numerically by increasing the association of proteins)

A key insight generated by this model was that *in vivo* the PTS proteins exist primarily in complexes with other PTS proteins; verifying the analytically-derived atypical response coefficient sums in group transfer pathways. It is difficult to imagine a more succinct approach to discovering these pathway behaviours. The modular nature of "bottom-up" models allows this PTS model to be included in larger bacterial models. However, when transferring a sub-model into a larger model, care needs to be taken to account for the particular conditions represented. For instance, maximal reaction velocities are very dependent upon enzyme concentration and temperature, factors which are determined by the particular growth and experimental conditions employed. And thus certain parameters (such as V_{max} values) have to be re-evaluated under alternate experimental conditions when necessary. Indeed, this will be illustrated by the inclusion of the PTS model developed by Rohwer *et al.* [68] in the glycolytic models developed in this dissertation. The concentration of the first protein component of this pathway, EIICB, is known to fluctuate up to ten-fold in response to changes in substrate availability [76, 77] (a flexibility which was exploited in Sections 5.6 and 5.7).

1.3 Generic rate equations: the generic reversible Hill equation

A requirement for systems modelling is a set of simple and versatile enzyme kinetic equations. It is often both experimentally unachievable and unnecessary to factor explicit reaction mechanisms into rate equations (e.g. mass-action kinetics). Rather it is often sufficient to derive phenomenological equations that describe the response of enzymic reactions to changes in substrate, product and modifier concentrations [78, 79]. Various attempts have been made to produce a valid description of these activities. For instance, a simple equation such as the uni-substrate irreversible Michaelis-Menten [80–82] is somewhat artificial in that it does not factor in reversibility, product-inhibition, cooperativity and multiplicity of substrates; traits that are important to consider when modelling biological pathways. A collection of equations has been derived which incorporate some of these traits, but most do not include at least one or are inconvenient to use due to a large number of parameters without clear operational definitions (e.g. the reversible Monod, Wyman and Changeux equation, MWC) [83]. Additionally, certain kinetic equations are not accurate over all possible biological conditions. The Adair equation was formulated to explain cooperative binding kinetics in haemoglobin but proved inadequate when applied to enzymes with more than two binding sites by producing contradictory cooperativity relationships between the various dissociation constants [84]. This shortcoming was due to an assumed independence of binding at different binding sites on an enzyme. The Adair model was superseded by the previously mentioned MWC model and the Koshland, Némethy and Filmer (or "Sequential") model, which was based on the assumption of induced-fit cooperativity between enzyme binding sites. Other models propose that association and dissociation of proteins producing various states of aggregation are responsible for the cooperativity phenomenon [85].

The Hill equation incorporates essentially Michaelis-Menten kinetics with "Hill coefficients" as exponents which modulate the cooperativity of the enzyme either positively or negatively without prior knowledge of enzyme mechanism [86]. This equation was modified by generalising it to be reversible and incorporating allosteric modifier effects for uni-substrate enzymes [83]. The Generic Reversible Hill Equation (GRHE) accurately models allosteric modifier behaviour and cooperativity and is expanded to cover multi-substrate reactions with multiple allosteric modifiers [79]. Moreover the simple operationally-defined kinetic parameters of the GRHE (e.g. $S_{0.5}$, is the substrate concentration at which the reaction performs at half the maximal rate) are accessible to experimental determination, unlike the parameters of the MWC equation which have not got clear operational definitions and thus cannot be independently determined.

A selection of the resulting equations is shown below [79, 83]:

Uni-substrate reversible Hill:

$$v = \frac{V_f \sigma \left(1 - \frac{\Gamma}{K_{eq}}\right)(\sigma + \pi)^{h-1}}{\frac{1+\mu^h}{1+\sigma^{2h}\mu^h} + (\sigma + \pi)^h} \tag{1.1}$$

Generic Reversible Hill equation: (multisubstrate, no allosteric modifier term):

$$v = V_f \prod_{i=1}^{n} \sigma_i \left(1 - \frac{\Gamma}{K_e q}\right) \prod_{i=1}^{n} \left(\frac{(\sigma_i + \pi_i)^{h-1}}{1 + (\sigma_i + \pi_i)^h}\right) \tag{1.2}$$

Bi-bi Generic Reversible Hill equation: (bi-substrate, with allosteric modifier terms)

$$v = \frac{V_f \sigma_1 \sigma_2 \left(1 - \frac{\Gamma}{K_{eq}}\right) (\sigma_1 + \pi_1)^{h-1} (\sigma_2 + \pi_2)^{h-1}}{\prod_{j=1}^{m} \left[\frac{(1+\mu_j^h)}{(1+\alpha_j^{4h}\mu_j^h)}\right] + \prod_{j=1}^{m} \left[\frac{(1+\alpha_j^{2h}\mu_j^h)}{(1+\alpha_j^{4h}\mu_j^h)}\right] + [(\sigma_1 + \pi_1)^h + (\sigma_2 + \pi_2)^h] + (\sigma_1 + \pi_1)^h (\sigma_2 + \pi_2)^h} \tag{1.3}$$

where the GRHE is expanded for n substrates and products and m modifiers, σ values are substrate concentrations scaled by their half-saturation constants ($\frac{s}{S_{0.5}}$), a parameter which is operationally similar to K_m in Michaelian kinetics; the π values are product concentrations scaled by their respective half-saturation constants ($\frac{p}{P_{0.5}}$); the mass-action ratio $\Gamma = \frac{\prod[products]}{\prod[substrates]}$ and the equilibrium constant K_{eq} are combined to create the thermodynamic term known as the disequilibrium ratio $\frac{\Gamma}{K_{eq}}$; h is the Hill coefficient describing binding cooperativity; μ parameters are allosteric modifier concentrations scaled by their half-saturation constants ($\frac{m}{M_{0.5}}$); and α defines a k-type interaction which modulates the effect that an allosteric modifier has on substrate- or product binding ($\alpha > 1$ indicates activation, $\alpha < 1$ inhibition).

One of the benefits of the equation as it stands in contrast to others (such as MWC) is the operational definitions of the various terms: the disequilibrium ratio (containing the thermodynamic information), the allosteric modifier term and the remaining kinetic properties are all separately defined and independently determinable. Reversible Hill equations have also been generated for asymmetric reactions including the uni-bi, bi-uni, bi-ter and ter-bi. Setting the Hill coefficient to 1 reduces these equations to their non-cooperative counterparts [78].

It has been demonstrated that the set of universal Hill equations exhibit realistic kinetics by comparing the non-cooperative generalised form to detailed mechanistic models and the cooperative form with allosteric modification to the MWC model using data from pyruvate kinase [79, 87]. In the compulsory-order ternary-complex enzyme mechanism ("ordered") substrates bind the enzyme in a specified order before catalysis and products are also released in order. The alternative substituted-enzyme mechanism ("ping-pong") involves modification of the enzyme by the first substrate which allows it to bind the second substrate once the first product has been released. Detailed kinetic equations describing these reaction mechanisms can be derived using the King-Altmann method [84]. *In silico* comparisons between the ordered and substituted equations and the non-cooperative generalised reversible Hill equation over a substrate range spanning two orders of magnitude to simulate a theoretical "worst-case scenario", show that the generic equation is able to mimic the behaviour of both the ordered and ping-pong mechanisms [79, 88].

Allosteric modification provides cooperative enzymes with the necessary regulation to be able to respond appropriately in physiological contexts. In contrast to the commonly used MWC equation for allosteric modification of cooperative enzymes, the reversible Hill equation accurately pre-

dicts modifier saturation which results in a loss of allosteric effects at high substrate concentrations (the generalised version of the MWC does model modifier saturation appropriately) [78, 83, 89]. The reversible bi-substrate Hill equation was shown to correctly predict modifier-saturation in experiments with *Bacillus stearothermophilus* pyruvate kinase using the inhibitor NaH_2PO_4 [87]. One advantage that the MWC has over the GRHE however, is that the MWC is able to account for complex allosteric relationships by virtue of its characterisation of an enzyme as in equilibrium between two different kinetic states; for example a situation in which modifiers have different effects at high and low concentrations (e.g. the interaction of pyruvate kinase with Ca^{2+} [90]).

A development with high utility that deserves mentions is that of "convenience kinetics" [91]. In this schema Michaelis-Menten kinetics are generalised to represent reversible reactions with any range of stoichiometries and allosteric regulators, by assuming a simple random-order reaction mechanism. Many of the kinetic parameters are amenable to experimental determination [92]. One of the key goals of convenience kinetics is to avoid the thermodynamic interdependency of kinetic parameters, which can hamper parameter fitting by virtue of the danger of violating the second law of thermodynamics as a consequence of the constraint imposed by the equilibrium constant on the parameters [93, 94]. This is achieved by defining thermodynamically independent parameters which are not related by the Haldane equation [91].

1.4 SBML: a language for models

The core units of systems biology consisting of integrated theories, models and experiments require a standard representation if they are to be accessible by other researchers. A series of problems arise in the systems biology community with regard to publication of models including: the obsolescence of previous versions of model simulators, the difficulty of integrating information from various analytical tools used in the construction of the model and the obvious difficulty of attempting to access, analyse and understand a model published in an unfamiliar format. In response to these needs the Systems Biology Markup Language (SBML) was developed [95] based on XML, the eXtensible Markup Language–a set of rules designed for electronic documents with emphasis on internet publication [96]. SBML provides a common "language" for models to be communicated between different software packages. Amongst other components, SBML models consist of:

Compartments	finite containers within which reactions take place, defined with a name and optional volume or position (within a different container) values
Species	chemical substances, participants in the reactions defined by name and initial amount
Reactions	mathematical descriptions of the transformation taking place between particular species with associated rate-laws
Parameters	symbolised quantities describing a facet of the system either locally or globally
Unit Definitions	the units to describe quantities in a model
Rules	mathematical expressions defining parameters or setting constraints upon a model which cannot be defined under "Reactions"

1.5 Metabolic control analysis

Metabolic Control Analysis (MCA) is a mathematical framework often used in conjunction with ODE models to study the sensitivity of steady-state variables (fluxes and metabolite concentrations) to fluctuations in system parameters and local enzyme rates. MCA was developed independently by two groups in the 1970s [11, 12]. The core realisation of MCA is that control of variables is shared to a lesser or greater extent by all of the parameters of the metabolic system and can be described by three mathematically connected coefficients which are dimensionless ratios [11, 14]. Elasticity Coefficients represent the apparent kinetic order of a reaction, and describe the sensitivity of a local reaction rate (v) to infinitesimal changes in the concentration of a connected metabolite (s, "metabolite elasticity") or a parameter (e.g. $s_{0.5}$, "parameter elasticity"). Elasticities are experimentally measurable quantities and are essentially scaled partial derivatives of a reaction toward a metabolite or local parameter [15].:

$$\varepsilon_s^v = \frac{\partial v_i / v_i}{\partial s_j / s_j} = \frac{\partial \ln v_i}{\partial \ln s_j} \tag{1.4}$$

Response Coefficients describe the response of a system variable (flux or metabolite concentration, y) to infinitesimal changes in a local parameter such as enzyme concentration (p), that is, the local effect of a parameter change echoes through all the components of the network which translates into a systemic effect, shifting the system to a new steady state. Deriving response coefficients represents a form of Parameter Sensitivity Analysis. Response coefficients take the following form:

$$R_p^y = \frac{\partial y_i / y_i}{\partial p_j / p_j} = \frac{\partial \ln y_i}{\partial \ln p_j} \tag{1.5}$$

Control Coefficients characterise the dependence of steady-state variables (i.e. flux or homeostatic metabolite concentration) on local reaction rates and are equivalent to the quotient of the

response coefficient and the elasticity coefficient (a relationship known as the combined response property: $C_v^y = R_p^y/\varepsilon_p^v$ [11], where y is a state variable, v is a local reaction rate, and p is a system parameter). Control coefficients define how much control a specific reaction exerts over a system variable and thus are experimentally measurable.

$$C_v^y = \frac{\partial y_i/y_i}{\partial v_j/v_j} = \frac{\partial \ln y_i}{\partial \ln v_j} \tag{1.6}$$

A number of important theorems exist which define the relationships between control coefficients, and relate the control coefficients to the local properties of individual reactions [14, 15], respectively the summation and connectivity theorems:

1. Summation properties:

Flux-control coefficients: $\sum_{i=1}^{n} C_i^{J_m} = 1$

Concentration-control coefficients: $\sum_{i=1}^{n} C_i^{S_m} = 0$

2. Connectivity properties:

Flux-control coefficients and elasticities: $\sum_{i=1}^{n} C_i^{J_m} \mathcal{E}_{S_j}^{v_i} = 0$

Concentration-control coefficients and elasticities:[1] $\sum_{i=1}^{n} C_i^{S_j} \mathcal{E}_{S_k}^{v_i} = -\delta_{jk}$

[1] δ_{jk} is The Kronecker Delta, a function of 2 variables: $\delta_{jk} = 1$ when $j = k$ and $\delta_{jk} = 0$ when $j \neq k$.

where i is a particular reaction, n is the total number of reactions, J_m is a particular system flux, S_j and S_k are variable metabolite pools.

Theorems that usefully describe control coefficients in terms of elasticities are described below in the context of Supply-Demand Analysis.

Supply-demand analysis

Supply-Demand Analysis (SDA) is a framework of thinking about the behaviour of metabolic networks which emerged out of microeconomics as a model of stock flow around an equilibrium [97]. Reactions are grouped into supply and demand blocks around a specific metabolite (or moiety-conserved cycle) and are analysed in terms of block response coefficients described as the sum of the set of partial response coefficients of constitutive reactions within the block [15, 98]. Additionally, rate equations can be used in which kinetic and thermodynamic regulatory constraints are distinguished, making the two effects clearly identifiable under a range of metabolite concentrations [99]. The supply and demand rates, representing either single enzymes or a block of reactions leading to and from a particular metabolite, can be graphed on the same pair of logarithmic axes with flux on the ordinate and metabolite concentration on the abscissa. This provides a useful

visualisation of activity around the steady state (intersection of the two graphs) known as a "rate characteristic" [99]. The elasticities of supply and demand, which are represented by the slopes of the tangents to the two functions at the steady-state point (where the two rate characteristic lines intersect), are an indication of the sensitivity of the respective block to perturbations in the concentration of the intermediate species [98]. As control coefficients can be expressed solely in terms of elasticities, supply and demand rate characteristics are useful for elucidating the division of flux control between the two blocks and the degree of control over metabolite concentration [15]:

$$C^J_{supply} = \frac{\varepsilon^{demand}_s}{\varepsilon^{demand}_s - \varepsilon^{supply}_s} \quad (1.7) \qquad C^s_{supply} = \frac{1}{\varepsilon^{demand}_s - \varepsilon^{supply}_s} \quad (1.9)$$

$$C^J_{demand} = \frac{-\varepsilon^{supply}_s}{\varepsilon^{demand}_s - \varepsilon^{supply}_s} \quad (1.8) \qquad C^s_{demand} = \frac{-1}{\varepsilon^{demand}_s - \varepsilon^{supply}_s} \quad (1.10)$$

where J is a steady state flux and s is a steady state metabolite linking the supply and demand blocks.

An important realisation of SDA and MCA is that control over flux is mutually exclusive (an implication of the flux control summation theorem: $\sum_{i=1}^{n} C_i^{J_m} = 1$) [98, 100]; if $-\varepsilon^{supply}_s >> \varepsilon^{demand}_s$ then C^J_{demand} tends towards 1 and assumes control over flux as C^J_{supply} tends towards 0, and *vice versa* [15]. Conversely, concentration control coefficients are inversely proportional to the sum $\varepsilon^{demand}_s - \varepsilon^{supply}_s$. Therefore, the greater the absolute block elasticities are, the more stringently homeostatic maintenance is imposed on the concentration of s and as one block elasticity tends to 0, assuming control over flux, the degree of homeostasis is inversely proportional to the elasticity of the other block. Considering a physiological concentration range for the metabolite of interest (sensibly extending up to near-equilibrium values where thermodynamic down-regulation dominates kinetics) illuminates regions of important allosteric regulatory phenomena such as those in which flux is controlled almost entirely by one particular block and important metabolites are homeostatically maintained over a narrow concentration range [98].

SDA has been applied to the analysis of control of glycolytic flux in *E. coli* using a series of *atp* mutants [101]. Lowering the $\frac{[ATP]}{[ADP]}$ ratio in the cell through artificially induced ATP-hydrolysis by increased expression of the F_1 domain of the membrane-bound H^+-ATP synthase under the control of a variable promoter, increased glycolytic flux. By dividing the metabolic reactions of the cell into supply and demand blocks around the $\frac{[ATP]}{[ADP]}$ ratio, it was shown that the demand block (and thus enzymes external to glycolysis that consume ATP) exerts $> 75\%$ of the control over glycolytic flux. Consequently, the glycolytic (or supply) pathway would be expected to exert a minority control over flux and the supply block elasticity is proportional to the degree of homeostatic maintenance of the energy charge of the cell.

1.6 Modelling glycolysis in *E. coli*

1.6.1 *E. coli* as a model organism

E. coli is a widely studied model organism making it ideal for modelling purposes. As of 2009 the EcoCyc database, dedicated primarily to strain K12 MG1655, contains all of the 4460 *E. coli* genes of which ~60% have experimentally defined functions [102]. Comparison to other genome annotations shows that by 2007 *E. coli* had a significant lead in the number of experimentally-determined gene functions: *Saccharomyces cerevisiae* = 37%, *Arabidopsis thaliana* = 5%, *Mus musculus* = 4%, *Drosophila melanogaster* = 4% [103]. The database also contains all the known metabolic pathways of the bacterium including 1326 metabolic reactions and 1362 compounds [102].

As the model organism for bacteria, understanding the regulation of central carbon metabolism in *E. coli* will be immensely informative.

1.6.2 Chassagnole and Peskov: two models dissected

Several metabolic models have been constructed for *E. coli* employing various strategies: Flux Balance Analysis [54, 58, 104–107], Metabolic Flux Analysis [51, 108–110], Metabolic Pathway Analysis [63], construction of a core model [111] and petri-nets [112]. It is clear from a cursory review of the literature that constraint-based models have almost entirely dominated the range of modelling strategies for the last 20 years [113].

For the purposes of background to this project, two salient modelling examples emerge. Chassagnole *et al.* [46] modelled the dynamic metabolic response of aerobically grown *E. coli* K-12 W3110 under glucose limitation to a glucose pulse in a chemostat at steady state using a constrained modelling strategy (an approach previously employed in [114]). Contrary to this approach, Peskov *et al.* [115] have produced a purely kinetic model more recently under almost identically aerobic physiological conditions. These are examples of two contemporary models of *E. coli* central carbon metabolism that incorporate a degree of kinetic information not present in purely constraint-based approaches. As such, they will be reviewed below.

The Chassagnole model

The strategy in adopted in [46] was essentially a hybridisation of constraint-based and purely kinetic strategies. The modelling process was begun by defining an exhaustive structural model accounting for the reaction stoichiometries of 107 metabolic pathways (702 individual metabolic reactions) available in the EcoCyc database at the time of publication [43]. Experimentally-derived uptake and excretion rates were used in conjunction with an estimation of the macromolecular composition (at a specified growth rate, μ) to solve a set of mass-balance equations and determine the steady-state fluxes. Metabolite concentrations were determined separately for a sub-second time course (stopped-flow technique [116]) and over 3-second intervals (quenching aliquots [117]) to monitor the immediate- and intermediate response to a glucose pulse respectively.

A set of kinetic equations and parameters was collected from the literature to constitute a starting model. Due to the danger of a multitude of unmodelled interacting processes inside the cell producing kinetics that are significantly different to the *in vitro* conditions in which enzyme kinetics are traditionally determined [46, 68, 114], the *in vivo* kinetic parameters (and some of the metabolite concentrations) were optimised by fitting to simulated dynamic data of intracellular metabolite concentrations using a subset of the steady-state ("stationary") model, called the "dynamic model". This submodel was restricted to glycolysis, the pentose-phosphate pathway (PPP) and the phosphotransferase system (PTS). An explicitly "top-down" structural decomposition technique was used in which kinetic parameters for the enzymes were estimated by progressively evaluating the network as a collection of sub-models forming metabolic "shells", each with an additional metabolite, beginning with the glucose uptake rate [114]. Unbalanced metabolites (those with neither source nor sink and thus no balance equations) were fitted with analytical functions to mimic dynamic measurements (including ATP, ADP, AMP, NAD, NADH, NADP, NAPDH). This stands in opposition to the "bottom-up" approach described above which seeks first to describe a metabolic network comprehensively by analysis and abstraction of elementary components and then validate the constructed model by experimentation; the result being a model which behaves realistically to induced changes in parameters or environment *in silico*.

The Chassagnole model fits the majority of the dynamic metabolite data after the glucose pulse with a few exceptions. At both the intermediate and immediate time spans (aliquots and stopped-flow extractions respectively) the model prediction of pyruvate levels showed the largest deviation from the data which could not be mitigated by changing the enzyme expression levels or optimising kinetic parameters of those reactions involved directly with the pyruvate pool.

The dynamic kinetic model was also subjected to metabolic control analysis to elucidate the distribution of control on glucose uptake by the PTS and to identify possible targets for redesign of the central carbon pathway. The highest glucose uptake Flux Control Coefficient (FCC) belonged to the PTS (0.42) with the next highest enzymes at 0.24 and 0.11, belonging to PFK and glucose 6-phosphate dehydrogenase (G6PDH) respectively; activities which are both able to indirectly regulate PTS activity by affecting the concentration of glucose 6-phosphate (G6P) which was modelled as feedback inhibitor of the PTS. The significant FCC for PFK is likely also due to the key allosteric terms included in the kinetic equation (PFK is modelled as being inhibited by PEP, a glycolytic product, and activated by ADP and AMP, which would signal a low cellular energy charge: $\frac{[ATP]+\frac{1}{2}[ADP]}{[ATP]+[ADP]+[AMP]}$). Pyruvate kinase which consumes PEP, the phosphoryl substrate for the PTS, only displayed a FCC of 0.01, indicating that it does not contribute significantly to the control of the rate of uptake of glucose despite having significant allosteric effectors included in its kinetic equation (inhibition by ATP, signalling high cellular energy charge; activated by FBP and AMP, signalling both the presence of substrate and low cellular energy charge).

This modelling approach may be industrially pragmatic but is not necessarily scientifically informative due to several caveats; some common to constraint-based modelling strategies. Firstly,

the kinetics of individual reactions are optimised to reproduce the observed rates and concentrations. This does not guarantee that the experimental data are accurate, or that the fitted parameter values offer a unique solution. Nor is it necessarily true that the model will behave realistically under different experimental conditions. Secondly, a number of kinetic parameters fitted with frankly unrealistic values (e.g. the GAPDH binding constant for bisphosphoglycerate = 0.0000104 mM) and were possibly unconstrained by the data (error estimates are not included). Additionally, a number of initial kinetic parameters were selected from data for different organisms. Lastly, the flux control coefficients calculated by the model rely heavily on the structure of the chosen kinetic equations. It is not clear in literature that either the concentration of G6P or the $\frac{[PEP]}{[PYR]}$ ratio are significantly responsible for regulation of the PTS (in contrast correlations between $\frac{[ATP]}{[ADP]}$ ratio and the proton motive force have been identified [70, 118, 119]), yet both are included in the simplified PTS rate equation and consequently figure prominently in the metabolic control analysis.

The Peskov model

The model developed in [115] is demonstrative of the design philosophy adopted in this project. It is a "bottom-up" kinetic model and was constructed by gleaning kinetic information from the literature and validated with published metabolomic and fluxomic data. The model consists of 75 reactions of 48 metabolites including the major central metabolic pathways: the PTS, glycolysis, the PPP, gluconeogenesis, the TCA cycle, the glyoxylate bypass and the Entner-Doudoroff pathway. These pathways were modelled as a series of ordinary differential equations (ODEs) each describing the change in a particular metabolite pool as the sum of reactions producing or depleting that particular pool (e.g. $\frac{dMet}{dt} = \Sigma v_{producing} - \Sigma v_{consuming}$). The model also included a number of conservation laws (e.g. $AXP_{TOT} = ATP + ADP + AMP + cAMP$) and simple kinetic rate laws to describe recycling (e.g. NADH $\rightarrow$ NAD^+) and export reactions.

To avoid an excess of free parameters (i.e. those without literature values), enzymic reactions were modelled with kinetic equations at 4 levels of mechanistic complexity. The first 3 levels were respectively allocated to enzymes that displayed low, intermediate, and highly complex kinetics and were modelled either according to the Cleland convention [120] or using the Monod-Wyman-Changeux equations [121]: low complexity reactions are represented by isomerisations and dehydrogenations which show no cooperativity or allosteric effects; those that were of intermediate complexity lacked any allosteric regulation but were complex enough to require data for verification; and lastly the highly complex reactions showed cooperativity and allosteric effects. The 4th degree of complexity was reserved for those reactions in which changes in enzyme state (e.g. protonation at different pHs) are modelled producing a separate ODE subsystem.

The vast majority of reactions (34 out of 44) were modelled using the two simpler models. Individual enzyme models (i.e. rate laws) were validated using *in vitro* initial rate assay data, whereas the combined metabolic model was verified using *in vivo* metabolomic and fluxomic data. A number of parameters were either fitted or deduced using the *in vivo* data, including maximal

rates, kinetic parameters of recycling reactions, and system outputs.

An interesting conclusion drawn from the modelling experiment was that at steady-state anaplerotic carbon is provided to the TCA cycle primarily through malate:quinone oxidoreductase (MQO) rather than either malate dehydrogenase (MDH) or the PEP/PYR bypass reactions (reviewed in [122]). Additionally it was shown that *in vivo* MDH flux was present, albeit to a lesser degree, in the opposite direction to MQO forming a futile NADH-consuming cycle (as demonstrated in [123]). The presence of a similar futile cycle *in vivo* has been shown to exist between PEP-carboxylase (PPC) and PEP-carboxykinase (PCK) [122, 124]. In the modelled flux, these reactions are in opposition and the PCK flux dominates possibly in response to an overproduction of oxaloacetate by the glyoxylate bypass pathway.

A key experiment performed during the modelling process succinctly demonstrates the power of modelling approaches to problem solving. The standard (or wildtype) model was modified to reflect a pyruvate kinase knockout (*PykA*) strain and compared to published metabolomic [125] and fluxomic [126, 127] data sets under the same conditions. The *PykA* model was able to reproduce the published fluxes but only upon introduction of a hypothesised allosteric inhibition of phosphogluconate dehydrogenase by PEP was the model equally able to reproduce the published steady state metabolite concentrations. Additionally the *PykA* model correctly predicted a compensatory increase in PPC flux [127].

A comparison of the Chassagnole and Peskov models clearly illustrates the differences between a "top-down" and "bottom-up" modelling philosophy and that one of the key strengths of a "bottom-up" approach is predictive power.

1.6.3 Metabolic strategies of *E. coli*: fermentation, aeration and efficiency

Mixed acid fermentation and microaerobic growth

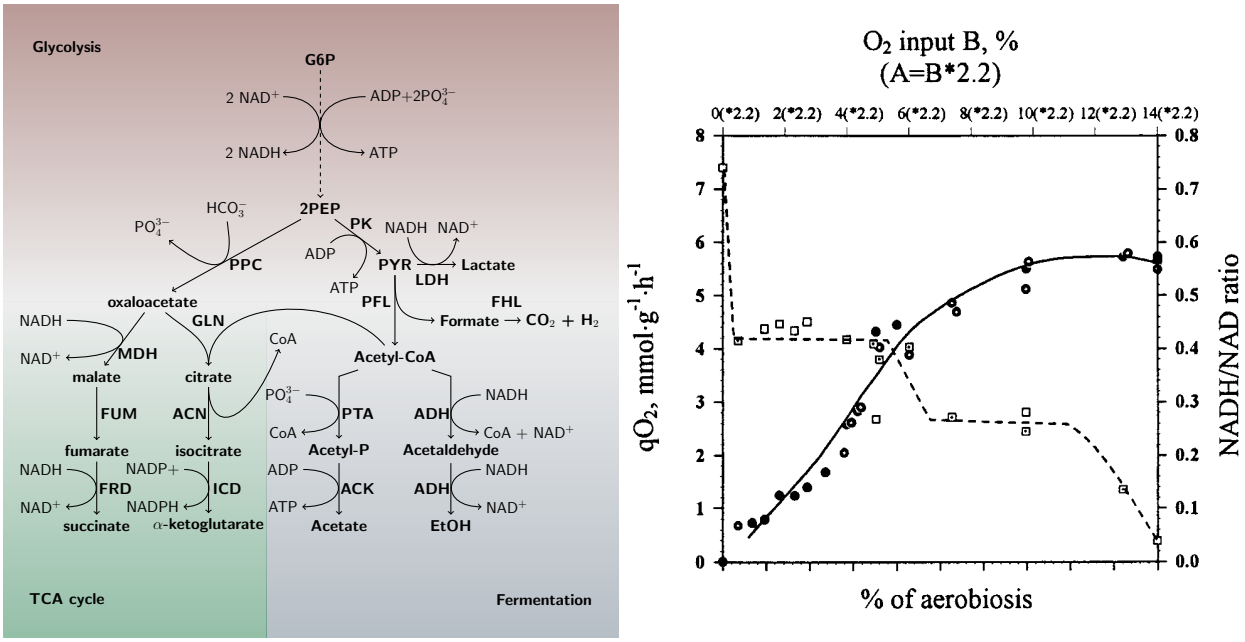
It is conceivable that in its natural environment *E. coli* rarely encounters purely aerobic or anaerobic conditions (e.g. O_2 partial pressure at the mucosal surface in the lumen of the small intestine is approximately 1/5th of that in the atmosphere [128]) and is additionally exposed to a variety of substrates. As such it is beneficial for the cell to be able to respond to these changing conditions with appropriate metabolic strategies. As a facultative anaerobe, *E. coli* has a broad spectrum of metabolic flux configurations for substrate-level phosphorylation in the absence of O_2; especially with regard to the anaerobic dissimilation of pyruvate (reviewed in [129]). Under O_2-limiting conditions *E. coli* cells employ a mixed acid fermentation, producing a diverse set of products, primarily comprised of lactate, ethanol, acetate, succinate, and formate (which is either exported from the cell at neutral pH or cleaved to H_2 and CO_2 at low pH and high formate concentrations) (Fig. 1.1a) [129, 130]. The specific ratio of fermentation products produced is modulated to maintain redox balance when growing on substrates with different oxidation states and in response to changing O_2 availability [129, 131, 132]. For example, the oxidation of glucose produces 2 NADH molecules (4 reducing equivalents, [H]) whereas the more oxidised gluconate and glucuronate (which are

metabolised via the Entner-Doudoroff pathway) produce only 1 NADH (2 [H]) each. Additionally, the first stage of glucuronate metabolism requires the input of one NADH as it passes through a fructuronate intermediate resulting in the redox-neutral production of two pyruvate molecules [129, 133]. Though conveniently requiring less redox balancing, these latter substrates have a lower energy yield as the Entner-Doudoroff pathway generates only one energy equivalent (ATP) per substrate molecule.

It is possible for redox balance to be achieved for all substrates of *E. coli* simply by varying the ratio of acetate and ethanol produced by fermentation, which has prompted questions about the use of the lactate and succinate generating pathways [129]. Succinate fermentation is a source of biosynthetic molecules and as such usually exhibits low flux [134], but has been shown to have the capacity for higher flux values in redirected *pykF* pyruvate kinase mutants [127]. Lactate fermentation however often dominates at pH values around and below 7.0 [129, 135]. Retaining the lactate fermentation pathway in evolution has been argued to be due to the possibility of generating an electrochemical proton gradient by efflux of lactate [136, 137].

Central carbon metabolism in *E. coli* is versatile and demonstrates a number of parallel pathways. Pyruvate, the end-product of glycolysis, can be cleaved into formate and acetyl-CoA by pyruvate-formate lyase (PFL) under typically anaerobic conditions or the pyruvate dehydrogenase complex (PDHC) under typically aerobic conditions [129]. The PDHC generates NADH and is for this reason primarily relegated to aerobic respiring conditions. A second branch point involves the consumption of acetyl-CoA by either the ethanol fermentation pathway which consumes 2 NADH, or the acetate pathway which produces 1 ATP but consumes no reducing equivalents. The final parallelism exists in the membrane-bound aerobic respiratory chain where electron flux is divided between two NADH dehydrogenases (NDH I which pumps two protons across the plasma membrane, and NDH II which is not coupled to proton-pumping) [138, 139] and two quinol oxidases ($cytb_o$ which pumps 2 protons across the membrane and $cytb_d$ which pumps 1 proton) [129]. $Cytb_d$ is a lower efficiency oxidase than $cytb_o$, but it has a higher affinity for O_2 making it a more efficient scavenger of O_2 under microaerobic conditions [132, 140, 141]. The inter-regulation of these four components gives the cell greater flexibility with regard to redox balancing [129, 142].

At a genetic level, the regulation of anaerobic, microaerobic and aerobic metabolism is achieved by two regulatory systems in *E. coli*: the ArcAB two-component system which regulates anaerobically- and aerobically-expressed genes, and FNR which is responsible for anaerobic regulation [129, 132, 143]. These two systems exercise an often inter-weaved co-regulatory strategy, neither being strictly responsible for a particular metabolic regime. For instance, as O_2 availability decreases ArcAB activates the expression of *cydAB* (encoding the high-affinity $cytb_d$), which is eventually inhibited by FNR once full anaerobiosis is achieved [132]. PFL, which predominates under anaerobiosis, employs a glycyl radical moiety and is thus irreversibly inhibited by O_2 [129]. This requires tight control of PFL activity by a PFL activase, PFL deactivase and possibly the scavenging of residual O_2 by $cytb_d$ [129, 132].



(a) (b)

Figure 1.1: Mixed acid fermentation and microaerobiosis in *E. coli*. (a) Alternate fates of glycolytic PEP during mixed acid fermentation in *E. coli*. At high glucose concentrations and under low O_2 availability succinate dehydrogenase, succinyl-CoA synthetase and α-ketoglutarate dehydrogenase are repressed producing a branched version of the TCA cycle (green) which produces biosynthetic precursors succinate and α-ketoglutarate. Mixed fermentation (blue) replenishes the NAD^+ pool for glycolysis (red) to continue. LDH - lactate dehydrogenase, PFL - pyruvate-formate lyase, PTA - phosphotransacetylase, ACK - acetate kinase, ADH - alcohol dehydrogenase, MDH - malate dehydrogenase, FUM - fumarase, FRD - fumarate reductase, GLN - citrate synthase, ACN - citrate hydro-lyase, ICD - isocitrate dehydrogenase, FHL - formate-hydrogen lyase, PO_4^{3-} - orthophosphate. (b) $\frac{NADH}{NAD^+}$ ratio (■ and □ represent two different chemostats) for growth of *E. coli* under various degrees of oxygenation. qO_2 - specific oxygen uptake rate (● and ○ represent two different chemostats); % aerobiosis is calculated from the amount of substrate converted to CO_2, assuming that at full aerobiosis 100% of the substrate is respired. Two distinct $\frac{NADH}{NAD}$ "plateaus" are observed and a transition point around 40 % aerobiosis. Sourced from [132]

In summary, the response of *E. coli* to varying levels of O_2 is characterised by two important adaptations that occur around the 40 % aerobiosis point [132]. Firstly, at high O_2 levels, *cydAB* is repressed and as O_2 levels drop ArcA (the cytoplasmic component of the ArcAB system) up-regulates *cydAB* expression to capitalise on the low O_2 levels. As O_2 availability decreases further, *FNR* binds the *cydAB* promoter to repress expression. Secondly, under aerobic conditions expression of *pflB* (encoding PFL) is minimal, but as O_2 availability decreases, *ArcA* and *FNR* bind to the *focA-pflB* promoter region to activate expression. Thus the switch is made from using the PDHC to PFL. In the 0 – 40% aerobiosis range (microaerobiosis) a "plateau" is observed in the $\frac{NADH}{NAD^+}$ ratio (= 0.4) which extends near to full anaerobiosis (Fig. 1.1b). Additionally, over this range (similarly at high glucose concentrations, Section 1.6.3) the full TCA cycle is inhibited and replaced by a branched version which produces succinyl-CoA by a reductive pathway and α-ketoglutarate by an oxidative pathway (Fig. 1.1a) [144]. Under low oxygen and high glucose conditions ATP is derived from glycolysis and the branched TCA cycle is split into a pair of biosynthetic pathways. Acetate, ethanol, and formate are produced in increasing proportions as fermentation gradually dominates respiration until full anaerobiosis is induced. At higher than 40 % O_2 availability another "plateau" is observed at a lower $\frac{NADH}{NAD^+}$ (0.25), the TCA is activated as a full cycle, the respiration rate increases, PDHC predominates over PFL, and $cytb_o$ takes over.

Efficiency vs. speed: the bacterial Crabtree effect

The tendency to select biomass maximisation as the objective criterion in constraint-based modelling risks inaccurate representation of the metabolism of many organisms that employ both high-efficiency/low-rate metabolic pathways such as aerobic respiration (producing ~30 ATP per glucose) and low-efficiency/high-rate fermentative pathways (producing 2-3 ATP per glucose) [145]. Moreover, low-efficiency pathways are often active even under conditions of abundant substrate and O_2; a phenomenon referred to in the literature as "overflow metabolism" or the "Crabtree effect". Examples of this phenomenon include amongst others: *Saccharomyces cerevisiae* which respires aerobically at low glucose concentrations and switches to primarily ethanol fermentation at high glucose concentrations while maintaining a minimal level of aerobic respiration; *Lactococcus lactis* which switches from mixed acid fermentation at low substrate concentrations to the low-yielding/high-rate homolactic fermentation at high substrate concentrations; and tumour cells which exhibit the "Warburg effect" and induce a combination of respiration and fermentation in the presence of substrate and abundant O_2 [123, 145, 146]. *E. coli* likewise in the presence of O_2 uses aerobic respiration at low glucose concentrations and switches to a lower-yield combination of respiration and "overflow" production of acetate at high glucose concentrations. A number of theories have been posited to explain this seemingly paradoxical observation. FBA models have been used to argue that when maximal capacities of the efficient (high-yield) pathways are reached, excess carbon flux "overflows" into ancillary pathways [54]. Besides being somewhat artificially imposed by virtue of the constraint-based modelling process [123, 145], this explanation does not account

for the fact that the transition at high substrate concentrations to inefficient pathways is regulated by the active suppression of efficient pathways in the organisms discussed [123]. Other theories include: a composite spatial/thermodynamic theory [147], and theories explaining the detrimental [148] and beneficial [149] effects of producing toxic by-products.

The unexpectedness of this behaviour has been largely dispelled by using a simple core-modelling approach which has shown that the switch to low-yield/high-rate fermentative metabolism at high substrate concentrations can be accounted for simply by the lower investment required to synthesise the protein constituents of the fermentative pathways [123]. In summary, unicellular organisms are competitively configured for fast replication, an advantage conferred upon them by the high-rates of substrate-level phosphorylation compared to respiration at high substrate concentrations. At decreasing substrate availabilities (and concomitantly decreasing growth rates) a threshold concentration is crossed beyond which it becomes too expensive in terms of substrate to maintain a high growth rate using the wasteful catalytically-efficient (low-yield/high-rate) pathway due to the input required to import substrate into the cell, and the cell switches to the metabolically-efficient (high-yield/low-rate) pathway and a general increase in glycolytic enzyme expression levels occurs [150].

Based on this model, under aerobic conditions and increasing growth rate the cellular redox balance governs the nature of the switch from high-yield/low-rate to low-rate/high-yield metabolism, that is, in organisms such as *S. cerevisiae* and *L. lactis* that don't express non-reducing alternative low-yield pathways for glycolytic flux, the switch between the two modes of metabolism is dramatic as one redox *milieu* is replaced by another [146]. This bimodality is not present in *E. coli* which exhibits a gradual and partial redirection of flux from aerobic metabolism to acetate in response to increased growth rate, maintaining simultaneously aerobic respiration and acetate production, capitalising on the increase in flux while minimising the penalty to metabolic efficiency via the extra ATP produced by the non-reducing acetate pathway [54, 146].

1.7 Nuclear magnetic resonance spectroscopy

NMR spectroscopy is a technique that exploits the behaviour of atomic nuclei in a magnetic field to make quantitative, dynamic and structural observations of chemical samples. Elementary particles (and thus certain atomic nuclei) exhibit a property called "spin", which is a quantised intrinsic angular momentum [151]. Isotopes with nuclei containing unfilled nuclear shells, and thus odd numbers of nucleons, exhibit a magnetic moment and angular momentum. Placing a sample of nuclei with angular momentum and random spin orientations in an external magnetic field (B_0) causes the spin orientations of the nuclei to align with the external magnetic field or exactly against it. The energy difference between nuclear spins in the low energy orientation (in the same direction as B_0) and the high energy orientation (against B_0) is described by equation 1.11 [152]:

$$\Delta E = \frac{\gamma h B_0}{2\pi} \tag{1.11}$$

where γ is the gyromagnetic ratio which relates the strength of the magnetic field (B_0) to the frequency of the nuclear spin, h is Planck's constant.

The ratio of nuclei in either orientation is quantified statistically by the Boltzmann equation [151]:

$$\frac{N_{upper}}{N_{lower}} = e^{-\Delta E/kT} = e^{-h\nu/kT} \tag{1.12}$$

where k is the Boltzmann constant and T is temperature.

The fundamental mechanism of NMR spectroscopy involves irradiating a partitioned distribution of nuclei in a magnetic field with a radio frequency (RF) pulse. A key property of the partitioning of spin populations is that the absorption of a photon of equal energy to the difference in energy between the spin states, causes spins to change orientation. According to the Bohr frequency condition:

$$\Delta E = h\nu \tag{1.13}$$

where ν is the resonance frequency, the frequency causing a spin transition can be written as:

$$\nu = \frac{\gamma B_0}{2\pi} \tag{1.14}$$

This frequency of radiation, produced in the experiment by an applied RF pulse, is known as the Lamor frequency. Absorption of RF energy causes the spin populations to equilibrate as low energy spins flip into the high energy orientation. Interpreting the spin population as a precessing net magnetisation vector (M_0) oriented in the direction of B_0 (the Z-axis according to convention), typical NMR experiments (as the ones used in this dissertation) involve pulsing the sample with RF radiation, which tips M_0 away from the state at rest, typically by 90 ° into the XY-plane. As the RF pulse is stopped, M_0 begins to precess around Z in the XY plane at the Lamor frequency, which can be observed using a detector in the NMR spectrometer and used to identify the particular nucleus. As molecular interactions create varying micro-environments for each nucleus, M_0 begins to return to equilibrium along the Z-axis, and the constituent nuclear magnetisation vectors begin to rotate out of phase by virtue of nuclei or "spin-packets" experiencing slightly different magnetic fields due to spatial displacement in B_0 [151]. These are known respectively as T_1 and T_2 relaxation processes and are responsible for the relaxation of M_0 to equilibrium.

The detection of the magnetisation vectors in a sample in the XY-plane produces a periodic signal known as a Free Induction Decay (FID) in the time domain. A key understanding related to the typical quantitative experiment performed in this dissertation is that identical nuclei in different molecular environments (e.g. a ^{31}P nucleus in two different molecules) exhibit different Lamor frequencies. The implication of this is that in a complex chemical sample the FID is composed of the signals of nuclei in many different environments, and Fourier transforming the FID from the

time domain to the frequency domain allows us to quantify each nucleus exhibiting an individual precession frequency. This is indeed one of the key benefits of NMR for studying complex enzyme reactions with cell extracts, that all metabolites are monitored simultaneously and directly, without coupling them to a downstream chromogenic reaction (as in the traditional spectrophotometric assays with NAD[P]H). Note therefore, that when an NMR spectrum is presented in this dissertation, the x-axis will be in the frequency domain (represented as parts per million of a reference frequency) which is positive on the left and negative on the right, that is, species to the left of another species rotate faster than their right-hand counterparts. The exact position of a nucleus on the x-axis is referred to as its "chemical shift" [151].

The ability of NMR to monitor most metabolites involved in a complex network of enzyme catalysed reactions makes it attractive for generating kinetic data. Indeed a wealth of information relating to reversibility, product inhibition, cooperative substrate binding, and allosteric modification can be gleaned from the data of NMR reaction time courses. In the course of this dissertation, two approaches to disentangling *in vitro* enzyme catalysed reaction time series data are presented: The first approach treats isolatable reactions or couples of reactions by fitting reaction time series with splines to approximate reaction rates (Chapter 3), the second approach treats networks of multiple inseparable reactions by fitting sub-models of the whole visible reaction network to the complex time series data (Chapter 4).

1.8 Justification

The current study seeks to integrate *in vitro* kinetic enzyme data from literature and a novel experimental framework into a kinetic model characterising the metabolic behaviour of glycolysis in *E. coli* W3110 under microaerobic conditions, by identifying *loci* of flux control and metabolite concentration homeostasis, with particular attention given to the dynamic behaviour of *E. coli* in response to varying degrees of O_2 availability and energy requirements.

A number of salient questions are approached in this study regarding the regulation of *E. coli* glycolysis including:

1. Is the "textbook" concept of a rate-limiting step in glycolysis valid for this model organism (usually PFK)?
2. Do ATP-demanding reactions external to glycolysis control flux through glycolysis in the microaerobic state (as shown for aerobic *E. coli* [101])?
3. Is the correlation between cellular free energy and PTS activity positive [118] or negative [101]?
4. Do any particular enzymes represent key regulatory points in glycolysis, especially with regard to the cellular free energy level and redox balance?

5. Are there structural features of glycolysis in *E. coli* which provide a unique degree of flexibility in response to redox demands?

6. To what degree is *E. coli* able to respond metabolically to slight variations in levels of O_2, and how much of the O_2 adaptation requires a transcriptional response?

7. Is it possible for FBP to act as an internal glycolytic flux-sensor (as suggested by Kochanowski *et al.* [153])?

A kinetic model of glycolysis and fermentation in *E. coli* provides a platform for answering these biological questions by *in silico* experimentation. The key regulatory points of the system are elucidated by MCA, and the behaviour of glycolysis in the context of external $\frac{[ATP]}{[ADP]}$ demand is quantified by SDA. Elementary Mode Analysis reveals the range of steady state flux behaviours achievable by the system. Ultimately, translating experimental hypotheses into perturbations to the system, which are only experimentally achievable by expensive and laborious means, can be performed rapidly and effortlessly *in silico*.

To achieve these goals the following chapters describe the strategy adopted in this dissertation and present the necessary techniques and research outcomes. Chapter 2 presents a software suite, NMRPy, developed for the efficient and accessible processing of NMR data. NMRPy proved useful as the key experimental techniques used in this study for the estimation of enzyme kinetic parameters utilised NMR reaction time series. Chapter 3 presents such a technique in which reaction time series are fitted with kinetic equations to estimate kinetic parameters. Chapter 4 describes the assembly of the glycolytic models and the model selection strategy. In Chapter 5 *E. coli* glycolytic behaviour is probed by various *in silico* experiments. And finally Chapter 6 discusses the development of these research techniques and their findings in the context of systems biology.

2 NMRPy: a Python module for the processing of nuclear magnetic resonance data

Summary

This chapter describes the development of a software toolkit using the Python programming language for the processing, analysis and visualisation of Nuclear Magnetic Resonance (NMR) data. Python is a high-level programming language which is eminently "readable" and supports a growing collection of "modules" produced in an open fashion, often collaboratively, by a community of developers for various applications, not least of which for scientific purposes. Though many open-source or freeware NMR processing software packages exist, very few are extensible, allowing users to customise the applications for their particular work. Additionally none provide a rapid and robust means of batch processing NMR spectra (and arrays of spectra) with a clear and easy-to-learn format. The NMRPy module was developed against a metabolomic and enzyme-kinetic background and is thus designed to work with 1-dimensional spectra and 2-dimensional time series arrays of spectra. Functionalities include amongst others: basic processing (zero-filling, apodisation, Fourier-transforming, baseline correction, phasing), deconvolution, wavelet transformation, b-spline approximation of time series integrals and visualisation of data (including interactive widgets for baseline correction, manual phasing, and peak picking). An introductory NMRPy tutorial is presented. NMRPy was used in this project and its use and usefulness is demonstrated in the following chapters.

2.1 Python and the open-source movement

One of the key requirements for this dissertation was an extensible software package which could be used for rapid batch-processing of large NMR datasets. It was found that none of the available NMR processing software suites was suitable. This is especially the case with regard to deconvolution of complex arrayed spectra, as produced by the NMR reaction time series in Chapters 3 and 4. Hence, an accessible and flexible NMR processing software was developed in the Python programming language which is made available to the public and is useful for many NMR processing applications.

Python (http://www.python.org/doc/) is a high-level dynamically typed, interpreted programming language which lacks the relative verbosity of other languages (e.g. C++, Java), providing a degree of readability which makes it ideal for dissemination of code in an 'open science' environment. Additionally, the comprehensiveness of Python's standard library (captured in its design motto: 'batteries included') and the openness and ease of development of modules for specific applications has earned Python a growing reputation in the scientific community as an extensible tool providing a platform to produce customised solutions to scientific problems. A cursory examination of the literature is sufficient to exhibit the wide range of scientific applications Python has been used for. A salient example is PySCeS [154], the Python Simulator for Cellular Systems which is written in Python and developed in our group to provide a platform for the construction, simulation, analysis and dissemination of ODE-based cellular models.

Recent decades have witnessed the development of a global trend which seeks the "opening up" of the pursuit of scientific knowledge by producing open-source software for scientific applications and making scientific data available for public utilisation and scrutiny, as well as providing a wealth of freely available online learning material [155]. Indeed this movement is not restricted to science and is eminently distilled in the development of major open-source software projects like the Linux operating system. With respect to software development, the transition to open-source has been described as the difference between the construction of a cathedral and a bazaar [156]: cathedral construction is a closed hierarchical process performed by a select group of specialists, whereas a bazaar emerges out of the coalescing of many people in public spaces with various skill-levels.

A vast number of open-source NMR-related packages are in use currently. These packages are tailored for specific categories of work: acquisition, analysis, processing, relaxation dynamics analysis and simulation, automated assignment, etc. With respect to data processing specifically, a number of packages appear on the NMR information website http://nmrwiki.org/:

- NMRPipe[157] — This popular collection of scripts and programs is powerful and has an exhaustive set of functions constructed around the UNIX pipe syntax, including: file format conversions and basic processing, peak-detection, Linear Prediction and Maximum Entropy Reconstruction, multidimensional (1D-4D) capabilities, and other high-end features for structure elucidation.

  [http://www.nmrscience.com/nmrpipe.html]

- SpinWorks — SpinWorks is a freeware and relatively lightweight application for the processing of 1D and 2D NMR spectra with a focus on organic and inorganic chemistry. In addition to basic processing functions, features include simulation and analysis of convoluted second-order spectra, and analysis of dynamic exchange spectra.

  [http://www.columbia.edu/cu/chemistry/groups/nmr/SpinWorks.html]

- PERCH NMR software[158] — Perch is a free package with a well developed GUI replete with processing, structure-prediction, quantification and visualisation features. Included are a number of advanced features for predicting spectra from 3D structures, simulating spin-spin systems and deconvolution of complex lineshapes.

  [http://new.perchsolutions.com/]

- Azara — A free command-line package with basic processing and peak-fitting capabilities.

  [http://www2.ccpn.ac.uk/azara/]

- GSim — A free GUI-based package for basic processing and integration of spectra from multiple sources.

  [http://sourceforge.net/projects/gsim/]

- MatNMR [159] — This package consists of a collection of Matlab scripts for the processing and analysis of NMR spectra (requires the commercial Matlab language http://www.mathworks.com/products/matlab/). MatNMR allows users direct access to numerical representations of NMR spectra. Features include basic processing functions such as linear prediction and arbitrary function apodisation.

  [http://matnmr.sourceforge.net/]

- ACD/Labs NMR software — This is a series of commercial applications made available freely for academic purposes. Basic processing functionality is included as well as peak-picking, integration and deconvolution capabilities. ACD software can process 1D and 2D NMR spectra, and arrayed 1D data.

  [http://www.acdlabs.com/]

- DOSYToolbox[160] — A simple but powerful package with a GUI for the processing and deconvolution of complex spectra with a focus on Diffusion-Ordered Spectroscopy.

  [http://dosytoolbox.chemistry.manchester.ac.uk/dosytoolbox/doku.php]

- NMRglue[161] — A Python module for the processing of NMR data from various sources. Functions primarily include basic processing and visualisation using the Matplotlib graphics library [162]. This package is as powerful as the user's skill with the Python programming language. Additionally it grants direct access to raw data in the form of Numpy [163] arrays.

  [https://code.google.com/p/nmrglue/]

- jMRUI[164] — A platform-independent software suite for the processing of NMR data. This is a powerful and highly accurate package which is designed for the exhaustive analysis of time-domain signals. To be usefully interactive we felt that the frequency domain was more useful for the current application as peaks could be readily identified and specified.

  [http://www.mrui.uab.es/]

A cursory examination of the available NMR software shows that packages are often developed with specific applications in mind and that the vast majority of software is designed for structure elucidation. Additionally there is a lack of appropriate non-commercial software for the rapid but rigorous processing of NMR time-series; of the packages mentioned above only the ACD/Labs package is able to rapidly process and quantify time series data, and it achieves this by using a crude block-integration method. As an extensible Python package NMRglue is more customisable than compiled gui-applications, however it is limited by the Python programming skill (or lack thereof) of the user and is not designed for time series data; as such it remains primarily useful for simple processing and 1D and 2D structure elucidation. Additionally NMRglue is still very much a work in progress.

NMRPy seeks to be a simple but powerful contribution to open science by making the processing of NMR spectra available to people with a broad range of skill-levels in various fields. It is an extensible solution for rapid NMR processing that requires only basic scientific computing skills for a standard NMR processing procedure. Though most functions pertinent to the processing, quantification, analysis and visualisation of 1D and arrayed 1D spectra are included in the NMRPy capabilities, the advanced Python user will easily be able to produce scripts for more complex applications as all data are accessible in Numpy arrays. Additionally, though NMRPy can be used in an interactive fashion using IPython, Python scripting provides an easy means of both designing scripts for a user's particular research application, and for batch processing large volumes of data with the option of parallel computing.

The following section 2.2 introduces a tutorial demonstrating a the power of NMRPy by following a typical workflow when processing NMR spectra. Thereafter, Section (2.3) will serve as a reference manual for NMRPy and a survey of its functionality. A list of commands and their descriptions is included under the headings:

2.4.1 Input/Output and Basic Processing

2.4.2 Peak-picking, Deconvolution and Integration

2.1.1 Acquiring NMRPy

NMRPy is available for download in a Git repository at https://github.com/jeicher/NMRPy.

2.2 Tutorial: an introduction to the use of NMRPy

This tutorial will demonstrate a comprehensive example of processing and visualisation of an NMR data array. The ^{31}P NMR spectra used in this example were accumulated at a frequency of 242.87 MHz on a Varian 600 MHz spectrometer with a 1 s repetition time (0.5 s acquisition/0.5 s relaxation). 100 transients were collected per individual FID using a pulse angle of 90° with no proton decoupling.

Typically NMRPy processing is performed in an interactive Ipython shell. However a script file containing standard processing commands and parameters can significantly reduce work time and provides an infrastructure for batch processing and parallel computing for large datasets. A simple script file will be assembled in this tutorial.
In the preamble of the script file the numerical and graphical modules are loaded:

```
import numpy as np
import matplotlib.pyplot as plt
```

Alternatively, NumPy and Matplotlib can be conveniently imported together as the combined namespace 'pylab':

```
from pylab import *
```

Next, the NMR processing package NMRglue including the data import function is imported, and finally NMRPy:

```
import nmrglue as ng
import nmrpy
```

2.2.1 Getting started: importing data

To import a FID (or FID array) the `ng.varian.read` function is used (`fid_directory` refers in this case to the Varian directory containing data, procpar and text files), after which a `FID` class is instantiated using the extracted data and parameter (`procpar`) files.

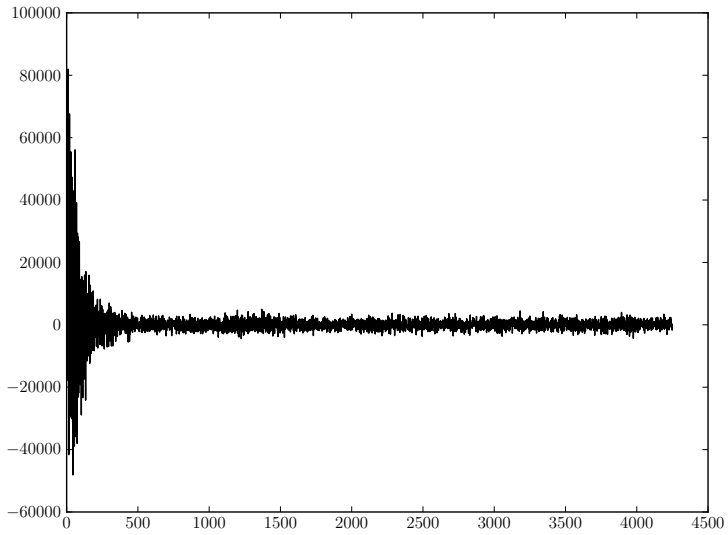


Figure 2.1: The raw FID data after import.

```
procpar,data = ng.varian.read(fid_directory)
p = nmrpy.f_fid(data,procpar,filename=fid_directory)
```

The `FID` object `p` is created and contains all relevant data, parameters and processing functions. The first spectrum of the raw data array looks as depicted in Fig. 2.1.

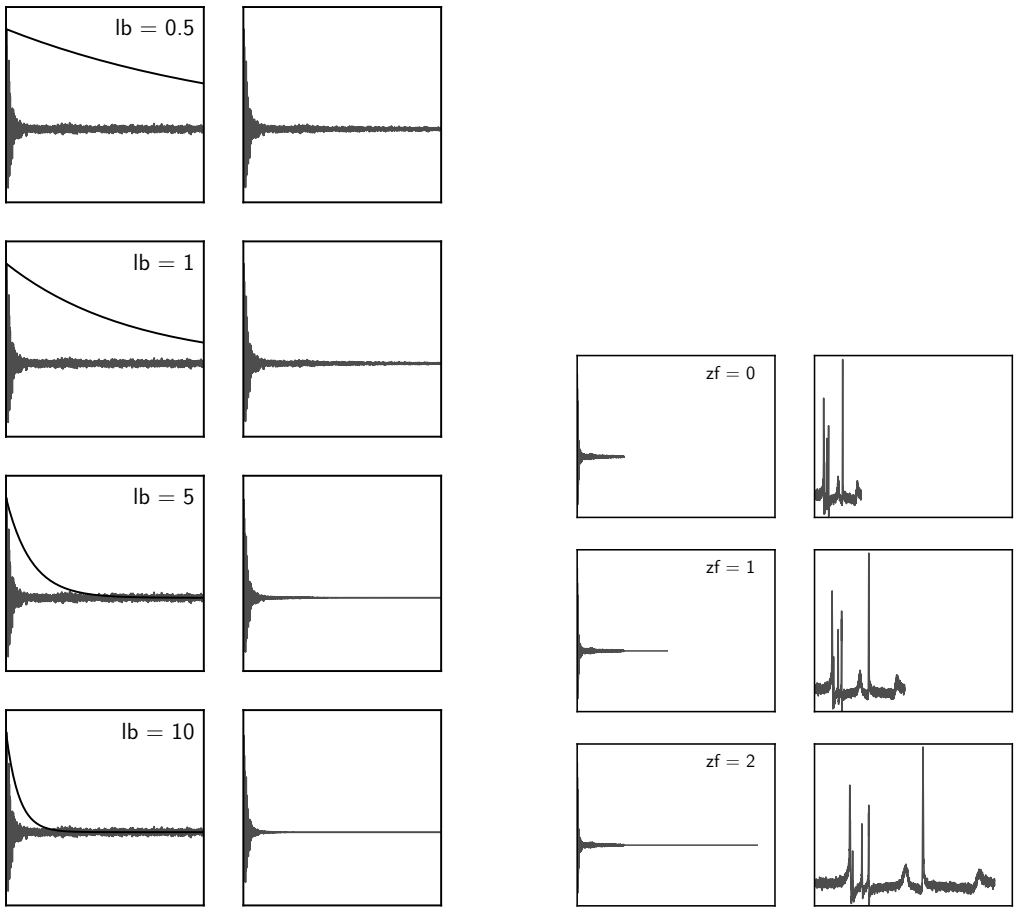
2.2.2 Basic processing: apodisation, zero-filling, phasing, baseline-correction

In this section basic processing procedures are going to be applied beginning with exponential apodisation (line-broadening) and zero-filling. Exponential apodisation removes noise acquired after the signal has decayed. The degree of apodisation is determined by several factors including the signal-to-noise ratio and the acquisition time. Fig. 2.2a illustrates the effect of apodisation at increasing degrees. In this tutorial a line-broadening value of 1 Hz will be used:

```
p.emhz(lb=1)
```

Zero-filling is a method of interpolating a FID before Fourier transformation by appending zeroes to the spectrum. This increases the digital resolution without additional information and results in narrower lineshapes with better definition. Fig. 2.2b illustrates the effect of zero-filling.
Two degrees of zero-filling are used in this example and the data is Fourier transformed. Individual spectra can now be visualised with the `plot_fid()` command (Fig. 2.3):

```
p.zf()
p.zf()
p.ft()
```



(a) (b)

Figure 2.2: (a) Exponential apodisation of a raw FID. Left blocks - the indicated apodisation function is superimposed onto raw data. Right blocks - the resulting apodised FID (b) Zero-filling of an FID. Left blocks - raw FIDs with indicated degree of zero-filling applied, Right blocks - respective Fourier-transformed spectra.

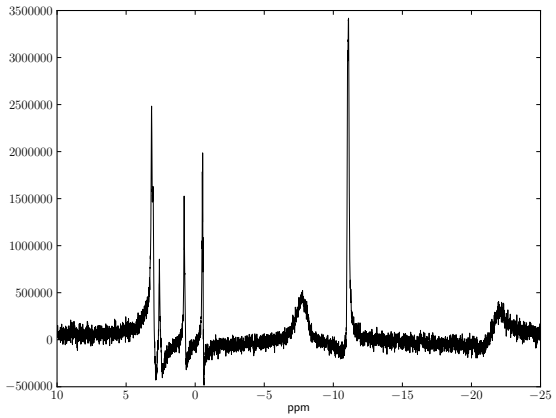


Figure 2.3: Fourier transform of an FID.

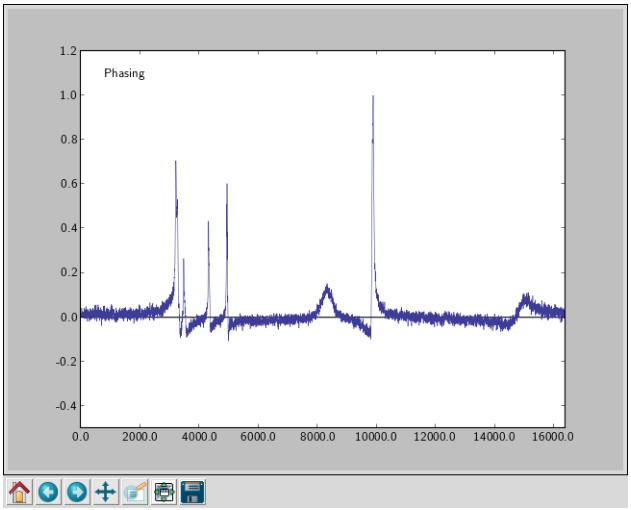
The following steps involve phasing and correcting baseline distortions in the spectra. Phasing can be performed manually or automatically. A manual phasing widget can be called using the `p.phase()` command (Fig. 2.4a). As a rough correction for large data arrays the manual phasing widget is called with the "universal=True" parameter indicating that the phase corrections of the first spectrum will be applied to all spectra in the array.
In this example automatic phasing is used with the objective criterion of minimising total area (other criteria are available, Sec. 2.4.1):

```
p.phase_area()
```

As the algorithm proceeds zero- and first-order phase corrections are output to the terminal (Fig. 2.4b) and the resulting array of phased spectra are stored (Fig. 2.4c).

Baseline corrections are performed by fitting a polynomial (of specified degree) to manually selected points on a spectrum. The point-selection widget is called (Fig. 2.5a) and a corrections are performed on each spectrum in the array by fitting a cubic polynomial to the selected points (Fig. 2.5b):

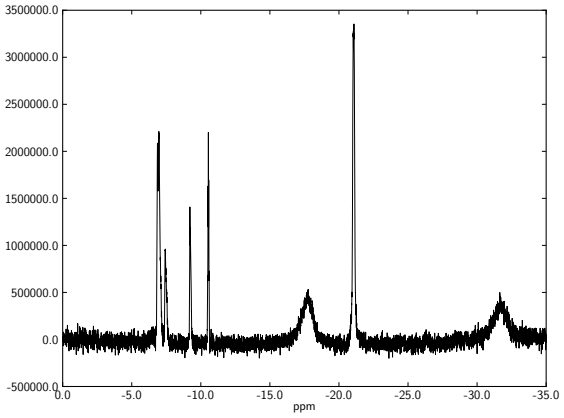
```
p.bl_select()
p.bl_fit(deg=3)
```



(a)

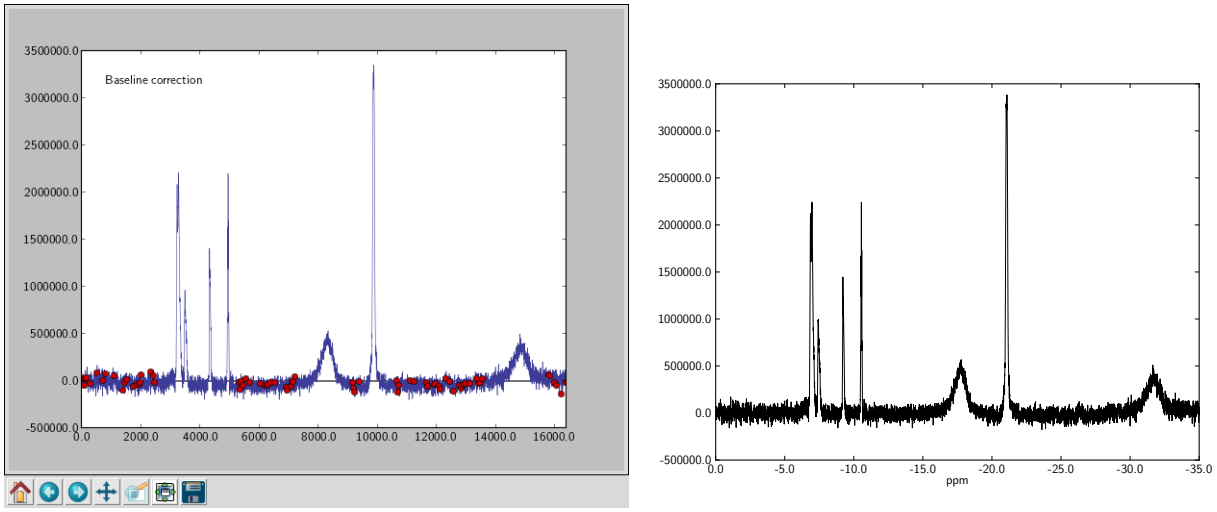
```
In [244]: p.phase_area()
0       p0: 73.9969878208       p1: -143.640336657
1       p0: 75.4794713605       p1: -146.049175981
2       p0: 74.5671010724       p1: -146.296910996
3       p0: 74.3136480032       p1: -144.780484486
4       p0: 75.3134083858       p1: -147.496667336
5       p0: 77.0586476646       p1: -150.0217446
6       p0: 75.0495943645       p1: -146.133783443
7       p0: 72.7605253181       p1: -143.616934409
8       p0: 75.7228022322       p1: -148.175007393
9       p0: 76.0328820411       p1: -148.711325037
10      p0: 74.8069537905       p1: -147.728262321
11      p0: 73.2754161812       p1: -144.325729061
12      p0: 75.5058681511       p1: -148.635036749
13      p0: 75.1918060065       p1: -145.900114861
14      p0: 76.1099734724       p1: -146.328212717
15      p0: 75.2238919247       p1: -144.996502449
16      p0: 74.3209531952       p1: -145.386372523
17      p0: 72.3740251588       p1: -140.244234224
18      p0: 76.3669052096       p1: -148.648081338
19      p0: 73.7537239963       p1: -146.735438002
20      p0: 75.1501426517       p1: -146.828835719
21      p0: 74.3583779827       p1: -144.59638386
```

(b)



(c)

Figure 2.4: Phasing of an FID.(a) Manual phasing widget (b) Terminal output of calculated phase corrections during autophasing (p0 - zero order, p1 - first order) (c) A phase-corrected spectrum. Y-axis is absolute intensity.



(a) (b)

Figure 2.5: Baseline-correction of a FID. (a) Point selection widget for baseline correction (b) Spectrum with baseline distortions removed. Y-axis is absolute intensity.

One of the particularly powerful features of NMRPy is the set of visualisation function for arrayed spectra, `plot_array()`. This function plots a clean 3D representation of the arrayed spectra and has numerous parameters allowing the user to produce literature-quality images (Sec. 2.2). In this example the up field boundary of the spectral width is specified (`=10`), the indices specify a range of spectra to plot (`[13,56]`, plotrange specifies the spectral width to plot (`[4.0,1.5]`), the individual spectra are filled (opaque polygons), and a dictionary of labels is provided. The resulting image is in figure 2.6.

```
labels =          {'FBP-A1':3.256,'G6P-B':2.843,'G6P-A':2.717,'FBP-B1':2.426,
                  'F6P':2.311,'FBP-B2':2.159,'FBP-A2':1.934}

p.plot_array_ppm(sw_left=10,index=[13,56],plotrange=[4.0,1.5],
                  filled=True,labels=labels)
```

2.2.3 Peak-picking, deconvolution and spline-fitting

What can be clearly seen in Fig. 2.6 is that the particular array being used in this tutorial involves a reaction that is changing over time. The next processing step is to integrate the spectra and quantify how much change is taking place. Due to the density and overlap of peaks in these spectra and the spectral drift a simple box integration function like `integrate()` would not be sufficient. NMRPy provides a deconvolution function `deconv()` which fits a template of peaks with specified spectral widths to the data. This template of peaks is composed of Lorentzian or Gaussian (or a combination of the two) lineshapes. `deconv()` thus requires a list of peaks and a list of fitting ranges

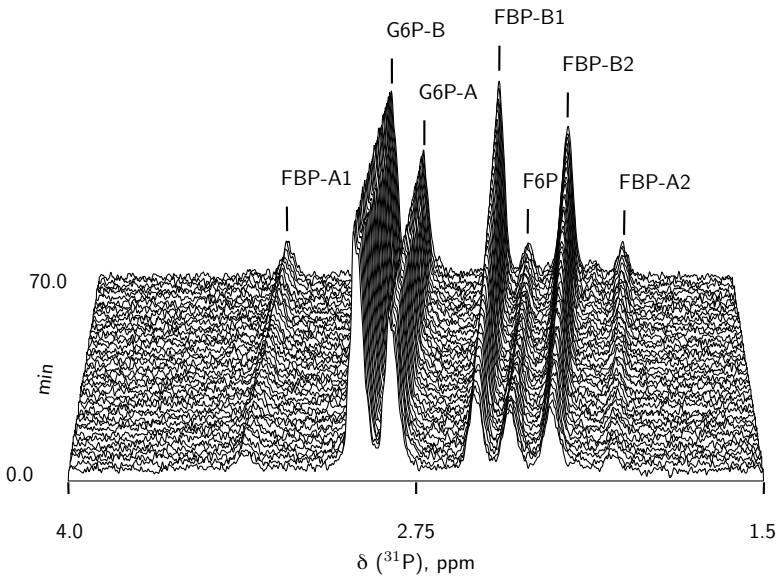


Figure 2.6: Array plot of processed spectra. FBP-A1/A2 – α-anomers of FBP, FBP-B1/B2 – β-anomers of FBP, G6P-A/B – α- and β-anomers of G6P

in a particular format. The algorithm then produces a template of lineshapes for each specified fitting range, and performs a convolution of this template with the spectral data to identify an ideal starting spectral shift with the highest overlap between the template and the data. After this, fitting proceeds using the Levenburg-Marquardt algorithm acting on several variables including, the spectral shift of the template peaks as a group, the spectral shifts of individual peaks within the group, the width of the peaks, and the amplitude of the peaks. The benefits of specifying fitting ranges are twofold: fragmenting large spectra into smaller individual datasets for fitting speeds up processing time significantly; specifying fitting ranges prevents lineshapes from "wandering" into alternate minima during the fitting process, an often encountered problem in highly convoluted spectra. The starting parameter lists can be easily generated by calling the `peakpicker()` widget (Fig. 2.7a):

```
p.peakpicker(index=20)
```

Peaks are selected by left-clicking the mouse and ranges are demarcated by right-clicking and dragging the pointer. Peaks appear as vertical red lines and ranges as light blue fields (figs. 2.7b, 2.7c and 2.7d). During this process the `tkinter` zooming and panning functionalities are available as buttons on the interactive window (Fig. 2.7a).

The resulting peak and range lists are as below. Note that each list is in fact a list of sequences; each fitting range list in `p.ranges` has an associated array of peak positions specified in `p.peaks` which fall within that range:

```
p.peaks = [array([ 3097.58672783,  3285.79052442,
                   3343.962607   ,  3487.68186985,
                   3542.43206522,  3614.29169664,
                   3713.52642575]),
           array([ 4913.64037806,  4925.71382916,
                   4936.6896938 ,  4950.95831783,
                   4961.93418247]) ]

p.ranges = [[2885.4297207667064, 3894.8864479185268],
            [4841.1996714467796, 5038.7652349315622]]
```

Deconvolution of an array of spectra is performed by fitting a series of lineshapes with default parameters generated using the `peaks` and `ranges` lists to each spectrum in the array:

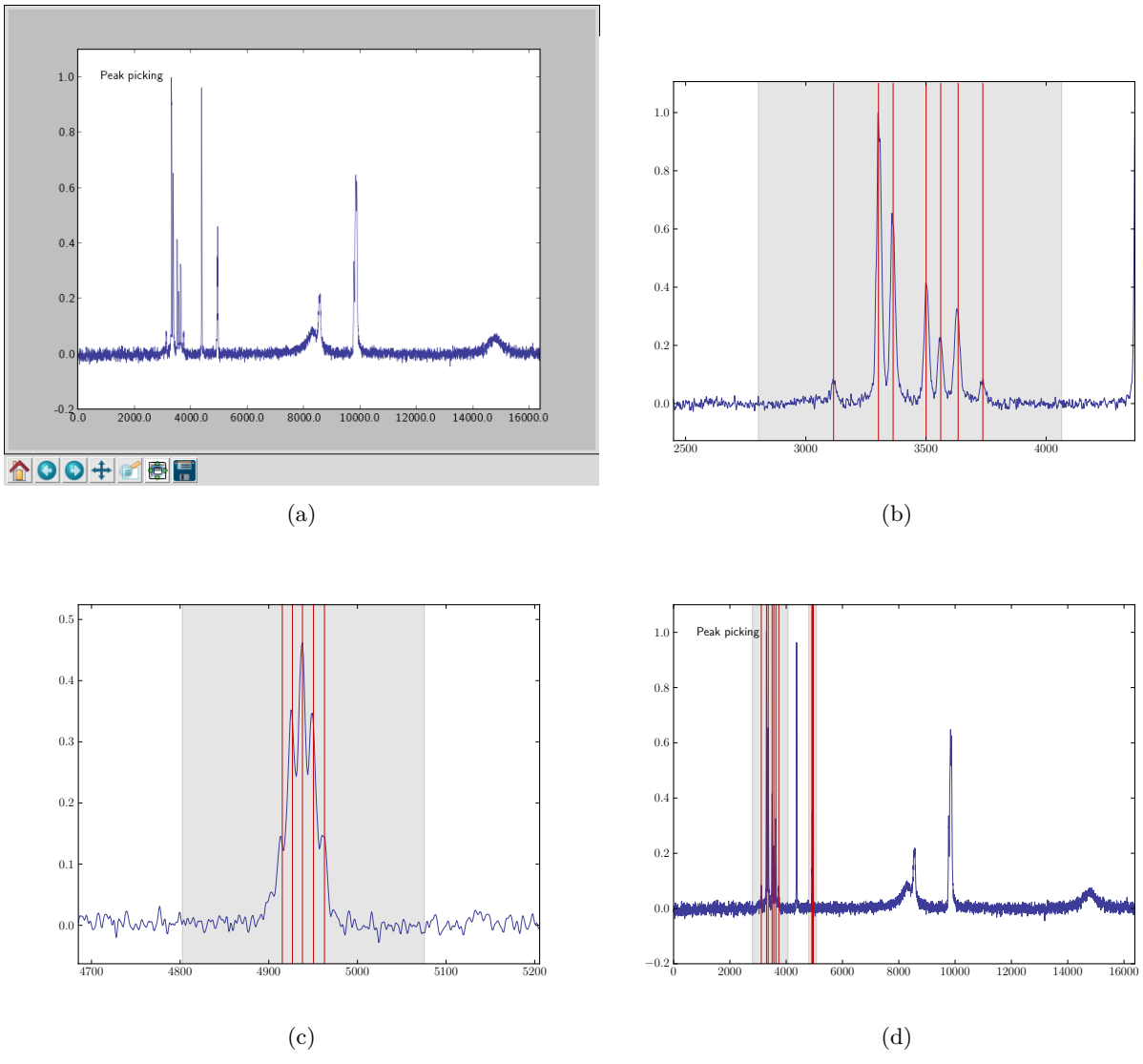
```
p.deconv(gl=0)
```

In this instance the parameter `gl` is set to 1 indicating a pure Gaussian peak (Sec. 2.2). As deconvolution proceeds, progress is indicated in the terminal (Fig. 2.8a). Fitted peak parameters are stored and can deconvoluted spectra can now be visualised using the `p.generate_plot( index=0 )` function as in Figs. 2.8b, 2.8c and 2.8d.

At this stage peak integrals for each spectrum have been stored in `p.integrals`. Summing J-coupled peaks or anomers of an individual species can be achieved using the `p.f_integral_sums( names, peak_index )` function, where `names` is a list of species names, and `peak_index` is a list of lists of peaks to sum:

```
names          = ['tep','g6p','f6p','fbp']
peak_index     = [[7,8,9,10,11],[1,2],[4],[0,3,5,6]]
```

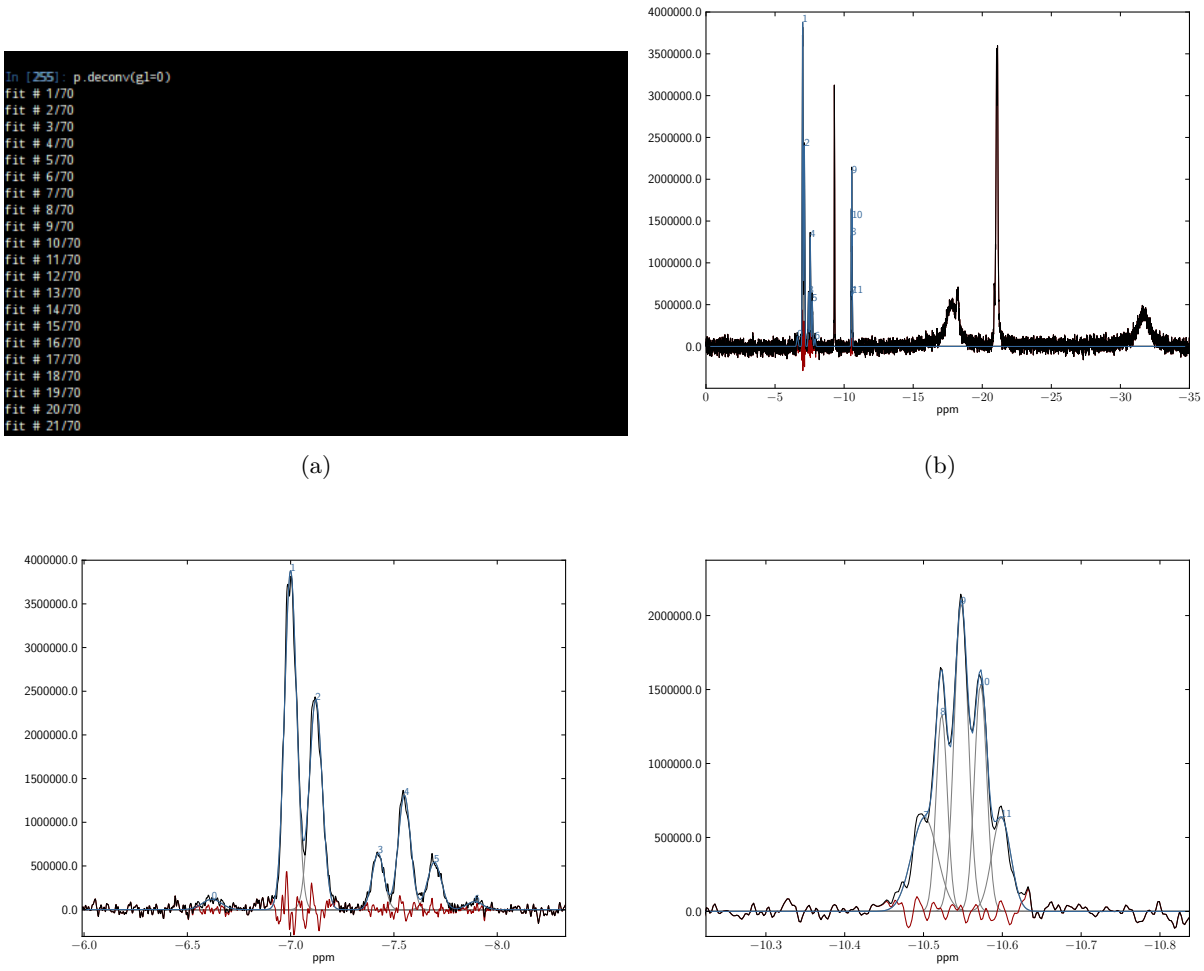
After this operation summed integrals are stored in an array `p.integrals_sum`. Triethylphosphate (TEP) was included at a concentration of 5 mM as an internal standard in this assay and as such will be used to determine the relative concentrations of the other metabolites:



(a) (b)

(c) (d)

Figure 2.7: Peak-picking. Red lines - peaks, blue fields - ranges. Y-axis is normalised intensity.



(a) (b)

(c) (d)

Figure 2.8: Deconvoluting a series of spectra. (a) terminal output during deconvolution, (b) - (d) deconvoluted spectra: black - data, blue - fitted lineshapes, red - residual. Y-axis is absolute intensity.

```
p.integrals_sum = p.integrals_sum/p.integrals_sum[0]*5.0
```

The scaled summed integrals can now be used by the spline-fitting (or polynomial-fitting) function to smooth time series data and extract rate approximations as b-splines are differentiable. In this instance `k`, the degree of spline to be fitted, is a list specifying the desired degree of spline-fit appropriate for each metabolite; `s` specifies the degree of smoothing to impose:

```
p.f_splines(k=[1,1,2,1],s=1000)
```

Fitted spline integrals (typically concentrations) and rates are stored as `p.s_concs` and `p.s_rates` respectively, and can be plotted using `p.f_plot_splines()` (Fig. 2.9):

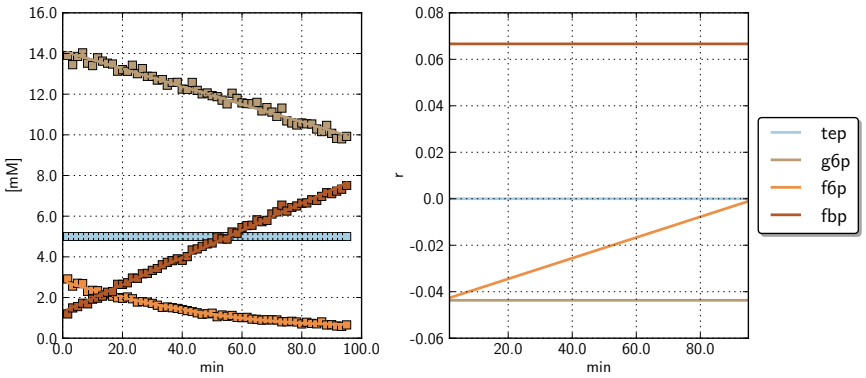


Figure 2.9: Spline-fitted time series data. Left - concentrations, Right - rates (mMol/min).

A full script is included below:

```
1  from pylab import *
2  import nmrglue as ng
3  import nmrpy
4
5  fid_directory = 'tutorial2.fid'
6  procpar,data = ng.varian.read(fid_directory)
7  p = nmrpy.f_fid(data,procpar,filename=fid_directory)
8
9  p.emhz(lb=1)
10 p.zf()
11 p.zf()
12 p.ft()
13
14 p.phase_area()
15 p.bl_select()
16 p.bl_fit(deg=3)
```

```
17
18 p.plot_fid(index=40)
19
20 labels = {'FBP-A1':3.256,'G6P-B':2.843,'G6P-A':2.717,'FBP-B1':2.426,'F6P':2.311,
21                 'FBP-B2':2.159,'FBP-A2':1.934}
22 p.plot_array_ppm(sw_left=10, index=[13,56], plotrange=[4.0,1.5], filled=True,
23                 labels=labels)
24
25 #p.peakpicker(index=20)
26 p.peaks = [ array([ 3097, 3285, 3343, 3487, 3542,  3614,  3713]),
27                 array([ 4913,  4925,  4936 ,  4950, 4961])]
28 p.ranges = [[2885, 3894], [4841, 5038]]
29
30 p.deconv(gl=1)
31
32 p.generate_plot(index=40)
33
34 names          = ['tep','g6p','f6p','fbp']
35 peak_index          = [[7,8,9,10,11],[1,2],[4],[0,3,5,6]]
36
37 p.f_integral_sums(names,peak_index)
38
39 #correcting for incomplete T1 relaxation
40 p.integrals_sum = map(multiply,p.integrals_sum,[6.51,5.14,5.04,2.61])
41 #scaling by internal standard
42 p.integrals_sum = p.integrals_sum/p.integrals_sum[0]*5.0
43
44 p.f_splines(k=[1,1,2,1],s=1000)
45 p.f_plot_splines()
```

2.3 NMRPy architecture

NMRPy was designed in an object-oriented fashion around a central 'unit': the 'FID' class. The FID class is instantiated at the beginning of processing and contains the data (1D spectrum or 2D array of spectra) and experimental parameters extracted from the associated Varian 'fid' and 'proc' files respectively; as well as a host of useful functions for processing, visualisation and analysis of the data. The NMRPy module has a number of dependencies:

SciPy and NumPy [163]	The Scientific Python and Numerical Python packages provide the numerical capabilities of NMRPy.
Matplotlib [162]	Matplotlib is a plotting library used for visualisations.
NMRglue [161]	NMRglue is a NMR data processing module for Python which provides the file import functionalities to NMRPy.
IPython [165]	IPython is an optional addition to NMRPy which provides an interactive python shell with a host of useful features for an interactive processing session (e.g. tab-completion, debugging, code history, support for interactive data visualisation, tools for parallel computing etc.)
PyWT [166]	An optional dependency, PyWavelets is a discrete wavelet transform library for Python, and provides a wavelet smoothing function for noisy data.

Typically a user will begin manual processing of a NMR spectrum or an array of spectra in interactive mode using IPython.

2.4 The FID class

2.4.1 Input/output and basic processing

`nmrglue.varian.read(dir='.', fid_file='fid', procpar_file='procpar', read_blockhead=False, shape=None, torder=None)`

This file import function is from the NMRglue package [161]. The following description is from the published man-page:

Read Varian files in a directory.

Parameters:

dir	*Directory holding Varian data.*
fid_file	*Name of fid file in directory.*
procpar_file	*Name of procpar file in directory.*
read_blockhead	*Set True to read blockheader(s). False otherwise.*
shape	*Shape of data, None tries to finds this automatically.*
torder	*Trace order (None for automatic). See read function.*

`f_fid( data, procpar, filename=None ):`

Instantiate the base FID class. Data must be a 1D or 2D NumPy array of spectra of equal size. Procpar is the Varian parameter file typically imported by nmrglue.varian.read().

`zf( self ):`

Apply a single degree of zero-filling resulting in a FID array extended in length to the next highest power of 2. Zero-filling is applied to the raw FID before Fourier transformation and serves to artificially increase the resolution of the data by interpolation.

`zf_2( self ):`

Apply a single degree of zero-filling resulting in a FID array of doubled length. Zero-filling is applied to the raw FID before Fourier transformation and serves to artificially increase the resolution of the data by interpolation.

`emhz( self, lb=10.0 ):`

Apply exponential apodisation function to spectra. The exponential decay function takes the following form:
$f(t) = e^{-\pi \frac{lb}{sw} t}$
where lb – degree of line broadening (Hz) and sw – spectral width (Hz). Exponential apodisation functions as a filter excluding noise from the raw FID before Fourier-transformation. This is necessary as acquisition times are typically longer than relaxation times and noise is acquired after the signal has decayed. Note that peak width at half maximum height (PWHM) in the Fourier-transformed spectrum $= \frac{k}{\pi}$, where the amplitude of the raw FID is described by e^{-kt}.
Keyword Arguments:
`lb` – degree of line-broadening in Hz.

`ft( self ):`

Fourier Transform the FID array. Calculates the Discrete Fourier Transform using the Fast Fourier Transform algorithm as implemented in NumPy [167]. Adapted from NMRglue [161].

`f_savefids( self, filename=None ):`

Save FID array to a binary file in NumPy '`.npy`' format.

`f_loadfids( self, filename=None ):`

Load FID array from a binary file in NumPy '`.npy`' format.

`phase( self, index=0, universal=False, norm=True ):`

Manual phasing of spectra. Instantiates a widget for manual phasing of specified FID.
Keyword arguments:
`index` – index of FID to phase
`universal` – phase a single spectrum and apply parameters to all spectra (will iterate over all spectra if `False`)
`norm` – normalise data
Note: left-click - phase p0, right-click - phase p1.

`phase_neg( self, thresh=0.0 ):`

Phase FID array automatically by minimising negative peaks. Uses the Levenberg-Marquardt least-squares algorithm [168] as implemented in scipy.optimize.
Keyword arguments:
`thresh` – threshold below which to consider data as signal and not noise (typically negative or 0)
Note: discards imaginary component.

`phase_area( self ):`

Phase FID array automatically by minimising area under peaks. Uses the Levenberg-Marquardt least squares algorithm [168] as implemented in scipy.optimize.
Note: discards imaginary component.

`phase_neg_area( self, thresh=0.0 ):`

Phase FID array automatically by minimising negative peaks and area under peaks. Uses the Levenberg-Marquardt least-squares algorithm [168] as implemented in scipy.optimize.
Keyword arguments:
`thresh` – threshold below which to consider data as signal and not noise (typically negative or 0)
Note: discards imaginary component.

`phasespace( self, index=0, inc=10 ):`

Evaluate total integral area of specified FID over phase range of p0: [-180,180] and p1: [-180,180] in degrees. Used for deriving rough phasing parameters for spectra that are difficult to phase.
Keyword arguments:
`index` – index of specified FID
`inc` – steps for range (-180 to 180 degrees) over which to evaluate
Returns a dictionary (`self.ph_space`) containing a 2D array of total integrated area of FID and associated phasing parameters.

`plot_phasespace( self ):` Produce a heat-map plot of the phase space evaluated for a specified FID generated by `self.phasepace()`. Minima are indicates on the map with a 'x'.

`bl_select( self, index=0 ):` Select points for baseline correction. Instantiates a widget to select points for baseline correction which are stored in `self.bl_points`.
Keyword arguments:
`index` – index of FID array to use for point selection.
Note: left-click selects point, right-click deselects point

`bl_fit( self, deg ):` Perform baseline correction by fitting specified baseline points (stored in `self.bl_points`) with a polynomial of specified degree (stored in `self.bl_polys`) and subtract these polynomials from the respective FIDs. Fitting is performed using the Levenberg-Marquardt least-squares algorithm [168].
Keyword arguments:
`deg` – degree of fitted polynomial

`f_wsmooth( self, level=20, thresh=2, filt='db8'):` Perform wavelet-smoothing on an array of spectra or a single spectrum. Uses the wavelet transform module PyWT.
Keyword Arguments:
`level` – decomposition level
`thresh` – smoothing threshold
`filt` – filter function

2.4.2 Peak-picking, deconvolution and integration

`peakpicker( self, index=0 ):` Pick peaks and integration ranges. Instantiate a peak-picker widget.
Keyword arguments:
`index` – index of FID array to use for peak selection.
Note: left-click - select a point, right-click - begin a drag-selection of a fitting range. Only peaks included in selected fitting ranges will be retained.

`integrate( self, index=None ):`

Perform a simple 'box integration' over a specified index using the ranges stored in `self.ranges`.
Keyword arguments:
`index` – indices of spectra to integrate, can be a single value or a list: `[low,high]`
Returns an array of integrals by spectrum.

`deconv( self, gl=0 ):`

Deconvolute an array of spectra (`self.data`) using specified peak positions (`self.peaks`) and ranges (`self.ranges`) by fitting the data with either a Gaussian, Lorentzian or a combined Gaussian/Lorentzian function. Uses the Levenberg-Marquardt least-squares algorithm [168] as implemented in scipy.optimize.
Keyword arguments:
`gl` – Gaussian fraction of peak function (0 – pure Lorentzian, 1 – pure Gaussian)

`f_integral_sums( self, names, peak_index ):`

Generate integrals for species by summing individual fitted NMR peaks (stored as `self.integrals`). Peaks are typically generated using `self.deconv()`.
Keyword arguments:
`names` – list of names of species e.g. `['species1', 'species2', 'species3']`
`peak_index` – list of peaks to sum per species e.g. `[[0,2], [1] ,[3,4,5]]`
Summed integrals are stored in `self.integrals_sum`.

2.4.3 Visualisation

`plot_fid( self, index=0, sw_left=0, lw=0.7, x_label='ppm', y_label=None, filename=None ):`

Plot an FID.
Keyword arguments:
`index` – index of FID array to plot
`sw_left` – upfield boundary of spectral width
`lw` – plot linewidth
`x_label` – x-axis label
`y_label` – y-axis label
`filename` – save file to filename (default `None` will not save)

`plot_array_ppm( self, plotrange=None, index=None, sw_left=None, lw=0.5, amp=0.5, azm=-90, elv=20, bh=0.05, filled=False, plotratio=[1,1], labels=None, fs=10, x_label=r'δ ($^{31}$P), ppm', y_label='min', y_space=None, filename=None ):`

Plot FID array in 3D using the Matplotlib mplot3d toolkit.
Keyword Arguments:
`index` – range of FID array to plot
`sw_left` – upfield boundary of spectral width
`lw` – linewidth of the FIDs in the plot
`amp` – normalised amplitude of the plot for scaling purposes
`azm` – azimuth of the plot
`elv` – elevation of the plot
`bh` – base box height in fraction of normalised data
`filled` – FIDs as filled polygons (True or False)
`plotratio` – list of XY scaling factors for plot window
`labels` – dictionary of peak labels and locations e.g. `{'spam': 5, 'eggs': -4}`
`fs` – label font size
`x_label` – x-axis label
`y_label` – y-axis label
`y_space` – y-axial displacement of x-axis label, defaults to: `-0.4acqtime[-1]` where `acqtime` is an array of acquisition times for each FID in minutes, representing the y-axis
`filename` – save file to filename (default `None` will not save)

`generate_plots( self, index=0, txt=True, sw_left=0, x_label='ppm', filename=None):`

Generate a plot of a specified spectrum with fitted peaks.
Keyword arguments:
`index` – index of data array to plot
`txt` – print peak number on plot (`True/False`)
`sw_left` – upfield boundary of the spectral width
`x_label` – x-axis label
`filename` – filename to save image under
Returns:
Plot of data (black), fitted Gaussian/Lorentzian peakshapes (blue), residual (red).

2.4.4 Spline-approximation of time series

`f_splines( self, s=1000, k=3):`	Use B-splines to approximate the summed integral time series (`self.integrals_sum`, generated by `self.f_integral_sums()`) as implemented in the scipy.interpolate module. Keyword arguments: `s` – smoothing factor of the splines `k` – order of the splines (can be a list to specify individual orders for species) `d` – order of derivative of the splines to compute (must be less than or equal to k; typically 1)
`f_polys( self, deg=5):`	Fit the summed integral time series (`self.integrals_sum`, generated by `self.f_integral_sums()`) with a polynomial of specified degree. Keyword arguments: `deg` – degree of polynomial to fit
`f_save( self, concs, rate, names, filename=None):`	Save spline-approximated concentration-rate data as text file. Keyword arguments: `concs` – concentration time series `rate` – rate data (as a single 1D array) `names` – names of species `filename` – filename to save data under
`f_plot_splines( self, x_label1='min', x_label2='min', y_label1='[mM]', y_label2='r'):`	Plot spline approximations of integral time-series generated using `self.f_splines()`. Left – integrals with spline approximations, right – derived rates.

2.4.5 Miscellaneous functions

These functions are typically only accessed by higher-level functions and are included for reference purposes.

`ps( data, p0=0.0, p1=0.0, inv=False ):`	Linear Phase Correction Keyword arguments: `data` – FID or array of FIDs `p0` – zero-order phase correction in degrees `p1` – first-order phase correction in degrees Adapted from NMRglue.
`f_pk( p, x ):`	Return the evaluation of a combined Gaussian/3-parameter Lorentzian function for deconvolution. Keyword arguments: `p` – parameter list: [`spectral offset (x)`, gauss: $2\sigma^2$, gauss: `amplitude`, lorentz: `scale (PWHM)`, lorentz: `amplitude`, fraction of function to be Gaussian `(0 -> 1)`] `x` – array of equal length to FID Note: specifying a Gaussian fraction of 0 will produce a pure Lorentzian and vice versa.
`f_pks( p, x ):`	Return the sum of a series of peak evaluations for deconvolution. See `self.f_pk() and self.deconv()`.
`f_res( p, data, gl ):`	Objective function for deconvolution. Returns summed error of the deconvolution fit. See `self.f_pk() and self.deconv()`.
`f_makep( data, peaks ):`	Make a set of initial peak parameters for deconvolution. Uses peaklist generated by `self.peakpicker()`. Keyword arguments: `data` – data to be fitted `peaks` – selected peak positions (see `self.peakpicker()` and `self.deconv()`)

`f_fitp( data, peaks, gl ):`	Fit a section of spectral data with a combination of Gaussian/Lorentzian peak for deconvolution. Keyword arguments: `data` – data to be fitted, 1D array `peaks` – selected peak positions (see `self.peakpicker()`) `gl` – fraction of fitted function to be Gaussian (1 - Guassian, 0 - Lorentzian) Note: peaks are fitted using the Levenberg-Marquardt algorithm [168] as implemented in scipy.optimize.
`f_conv( p, data ):`	Returns the translation of the maximum value of a convolution of an initial set of lineshapes and the data to be fitted. Used to determine ideal starting translation for fitting. Keyword arguments: `p` – parameter list: [`spectral offset (x)`, gauss: $2\sigma^2$, gauss: `amplitude`, lorentz: `scale (PWHM)`, lorentz: `amplitude`, fraction of function to be Gaussian `(0 -> 1)`] `data` – data to be fitted
`f_integrals( data, params ):`	Returns the integrals of a series of lineshapes for a single spectrum. Keyword arguments: `data` – data to be fitted `params` – fitted peak parameters
`f_integrals_array( data_array, param_array ):`	Returns the integrals of a series of lineshapes for an array of spectra. Keyword arguments: `data_array` – array of data to be fitted `param_array` – array fitted peak parameters

3 Determining enzyme kinetics for systems biology with nuclear magnetic resonance spectroscopy

This chapter has been published in the journal *Metabolites* [169] and is reproduced verbatim.

Summary

Enzyme kinetics for systems biology should ideally yield information about the enzyme's activity under *in vivo* conditions, including such reaction features as substrate cooperativity, reversibility and allostericity, and be applicable to enzymatic reactions with multiple substrates. A large body of enzyme-kinetic data in the literature is based on the uni-substrate Michaelis-Menten equation, which makes unnatural assumptions about enzymatic reactions (e.g. irreversibility), and its application in systems biology models is therefore limited. To overcome this limitation, we have utilised NMR time-course data in a combined theoretical and experimental approach to parameterise the generic reversible Hill equation, which is capable of describing enzymatic reactions in terms of all the properties mentioned above and has fewer parameters than detailed mechanistic kinetic equations; these parameters are moreover defined operationally. Traditionally, enzyme kinetic data have been obtained from initial-rate studies, often using assays coupled to NAD(P)H-producing or -consuming reactions. However, these assays are very labour-intensive, especially for detailed characterisation of multi-substrate reactions. We here present a cost-effective and relatively rapid method for obtaining enzyme-kinetic parameters from metabolite time-course data generated using NMR spectroscopy. The method requires fewer runs than traditional initial-rate studies and yields more information per experiment, as whole time-courses are analysed and used for parameter fitting. Additionally, this approach allows real-time simultaneous quantification of all metabolites present in the assay system (including products and allosteric modifiers), which demonstrates the superiority of NMR over traditional spectrophotometric coupled enzyme assays. The methodology presented is applied to the elucidation of kinetic parameters for two coupled glycolytic enzymes from *Escherichia coli* (phosphoglucose isomerase and phosphofructokinase). ^{31}P-NMR time-course data were collected by incubating cell extracts with substrates, products and modifiers at different initial concentrations. NMR kinetic data were subsequently processed using a custom software module written in the Python programming language, and globally fitted to appropriately modified Hill equations.

3.1 Introduction

A key requirement for the "bottom-up" approach to systems biology in which emergent systemic properties are inferred from a collection of basis elements, is accurate and comprehensive kinetic data; which, despite the existence of curated enzyme kinetics databases (e.g. BRENDA [170], SABIO-RK [171]), are often unavailable or inadequate for the desired experimental conditions. Experimental derivation of kinetic parameters can be expensive, labour-intensive, and often either overly simplistic and unable to comprehensively characterise enzymatic behaviour, or overly complex, having degrees of freedom that are beyond the dimensionality of experimental data [83]. Certain reaction characteristics such as reversibility and product-inhibition, and cooperative binding, which can be crucial to an *in vivo* understanding of a particular enzyme network, are at times dispensed with due to the paucity of experimental data [83]. Thus there is a need for an experimental system that is accessible and generates *in vitro* kinetic data to model the *in vivo* behaviour of enzymatic reactions comprehensively and accurately.

An additional requirement for systems modelling is a set of simple and versatile enzyme kinetic equations. The goal of enzyme kinetic modelling has traditionally been to elucidate and represent the detailed mechanisms of enzyme-catalysed reactions, often resulting in complex kinetic equations with numerous parameters [83]. Alternatively, in an effort towards simplification, unnatural assumptions are made that often result in arbitrary parameters without a clear operational meaning [83]. The Generic Reversible Hill Equation (GRHE) overcomes these obstacles by representing cooperativity, reversibility and allosteric behaviour with a minimal set of operationally-defined parameters, making it ideal for modelling of *in vivo* biological systems [79, 83]. Moreover, the kinetic parameters of the GRHE are amenable to direct experimental determination. For instance, the GRHE includes simple half-saturation terms for substrate, product and effector binding; an h value representing cooperativity of binding ($h > 1$ indicates positive cooperativity, $h < 1$ negative cooperativity, and $h = 1$ absence of cooperativity); and a modifier effect parameter, α, which determines the degree of positive ($\alpha > 1$) or negative ($\alpha < 1$) effect of the allosteric modifier on the reaction [79, 83].

Classical continuous enzyme assays involve collecting initial-rate data for a particular enzyme at various substrate concentrations and fitting these data to simple irreversible kinetic equations like the famous Michaelis-Menten [80–82] or Hill [86] equations. Less common is the alternative approach involving global fitting of complete progress curves of enzyme-catalysed reactions, instead of extracting initial rates from the first few data points and discarding the remainder of the time course [172]. This strategy involves either integration of the kinetic equation, making the substrate (and product) concentrations implicit, or differentiation of the time course data to generate rate approximations. The earliest attempts at progress-curve analysis involved the use of an integrated Michaelis-Menten equation fitted to simple single-substrate progress curves [173–175]. A more recent development is the closed-form solution of the integrated Michaelis-Menten equation using the Lambert-W function, which has been employed successfully for progress-curve analysis [176,

177].

The progress-curve strategy circumvents some of the issues of traditional initial-rate enzyme kinetics, such as burst-/lag-phases altering initial velocity measurements, experimental artefacts due to coupled enzymes, and the large number of experiments that need to be performed to generate relatively little kinetic data [178]. In contrast, progress curves are acquired while the substrate and product concentrations (and possibly also the effector concentrations) are changing, therefore yielding a relatively large amount of data per experiment on the substrate and product dependence of the reaction rate as well as on interference by inhibitors [174]. However, progress curve studies can be complicated by enzyme instability, which would augment perceived product inhibition and make the analysis of time course data significantly more complex [179].

Initial-rate kinetic assays are classically performed by coupling the reaction of interest via intermediate enzymes to a downstream chromogenic reaction (e.g. oxidation/reduction of NAD[P]H and NAD[P]$^+$) and monitoring the increase or decrease of the chromogenic substrate using a spectrophotometer. Initial rates are approximated by fitting a tangent line to the first few data points of the time course. A highly-sensitive discontinuous variant on this protocol involves downstream coupling to a cyclical pseudo zero-order reaction and approximating substrate and product concentrations by measuring the change in cycling rate. This sensitive and relatively laborious approach has been up-scaled and mechanised using a robotic platform [180]. Alternatives to these approaches usually involve a form of labelling (e.g. radiometric labelling [181]) or chromatographic techniques (e.g. HPLC, LC-MS [182]).

Unlike NMR spectroscopy, the approaches above are all labour-intensive, materials-intensive and are unable to provide direct real-time simultaneous quantification of substrate, product and effector concentrations. Moreover, recent improvements in the sensitivity of NMR spectroscopy show it to be an effective alternative for determination of enzyme kinetics that has been used successfully in conjunction with progress-curve analysis (e.g. assaying invertase and germacrene-D synthase [17]). ^{31}P-NMR involves a sufficiently sensitive NMR-active nucleus having an almost total natural abundance and relatively high gyromagnetic ratio, making this technique ideal for studying phosphorylated central carbon metabolites (^{13}C nuclei suffer from very low natural abundance and can only be employed if costly labelled metabolites are acquired, limiting the application of ^{13}C-NMR to this approach) [152]. An attractive feature of NMR spectroscopy is its applicability to *in vivo* metabolite measurements. This is a developing application, which is beginning to overcome the handicap of low sensitivity in whole organism studies (for example through the transfer of electron spin-polarisation to the nucleus of interest [18, 19]).

In this study, NMR progress curves of the two initial glycolytic reactions, phosphoglucose isomerase (PGI, EC 5.3.1.9) and phosphofructokinase (PFK, EC 2.7.1.11) in *E. coli* are acquired by incubating log-phase cell extracts with varying concentrations of substrate, product and effector. Generic Hill equations are parameterised by fitting to aggregated progress curves using a combination of genetic and least-squares algorithms.

E. coli PGI catalyses the first step of glycolysis after glucose transport via the PEP:glycose phosphotransferase system, is involved in gluconeogenesis, and serves as the branch point for entry into the pentose-phosphate pathway. PGI^- mutants reroute flux through the pentose-phosphate pathway and exhibit markedly decreased growth rates [183]. PGI catalyses the interconversion of glucose 6-phosphate and fructose 6-phosphate and exists in two forms: the major species making up more than 90% of the activity and consisting of two subunits, the minor being a dimer of the major species [184]. PGI is derepressed under anaerobic/micro-aerobic conditions [184] and has been shown to exhibit increased activity on a shift from an aerobic to a micro-aerobic environment [150].

E. coli has two phosphofructokinases, PFK-1 and PFK-2, which catalyse the conversion of fructose 6-phosphate and ATP to fructose 1,6-bisphosphate and ADP [129]. The primary enzyme responsible for $\approx$ 90 % of the total activity is the tetrameric PFK-1 [185]. The allosteric relationships of PFK are complex. The PFK reaction is moderately activated and inhibited under different conditions by its product ADP and other nucleoside di- and mono-phosphates, and inhibited by PEP, ATP and citrate [186–189]. At low F6P concentrations ADP activates the reaction and reduces cooperativity; at high concentrations of F6P, ADP inhibits PFK competitively with respect to ATP (product inhibition) and non-competitively with respect to F6P [187]. PFK is strongly inhibited by phosphoenolpyruvate in a glycolytic feedback loop [187]. PFK also exhibits a degree of positive cooperative binding towards F6P, and a negative cooperativity between ATP and F6P as well as between PEP and the substrates ATP and F6P [187, 189]. Many attempts at modelling the complex kinetics of PFK have been made, often employing the Monod-Wyman-Changeux model [187, 190]. It is clear that arriving at a comprehensive model of PFK kinetics is no mean feat.

3.2 Results and discussion

3.2.1 Method outline and technical considerations

Method outline

A method is presented by which kinetic equations are parameterised using NMR progress curve data for systems biological modelling (for a diagram illustrating the work flow in this study see Fig. 3.1): assays are prepared by combining various starting concentrations of substrate, product and effector (including all necessary co-factors and constituents necessary for NMR spectrometry) with a cell extract containing the enzyme(s) of interest. An appropriate NMR nucleus is selected and time courses are acquired as the enzymatic reaction(s) progress towards equilibrium. Each time course is captured as an array of Free Induction Decays (FIDs), which is processed and quantified to yield a standard progress curve of concentration change over time. Progress curves are fitted independently (i.e. not assuming mass conservation) with splines to approximate the reaction rate by differentiating the splines at each of the measured metabolite concentrations. The spline data

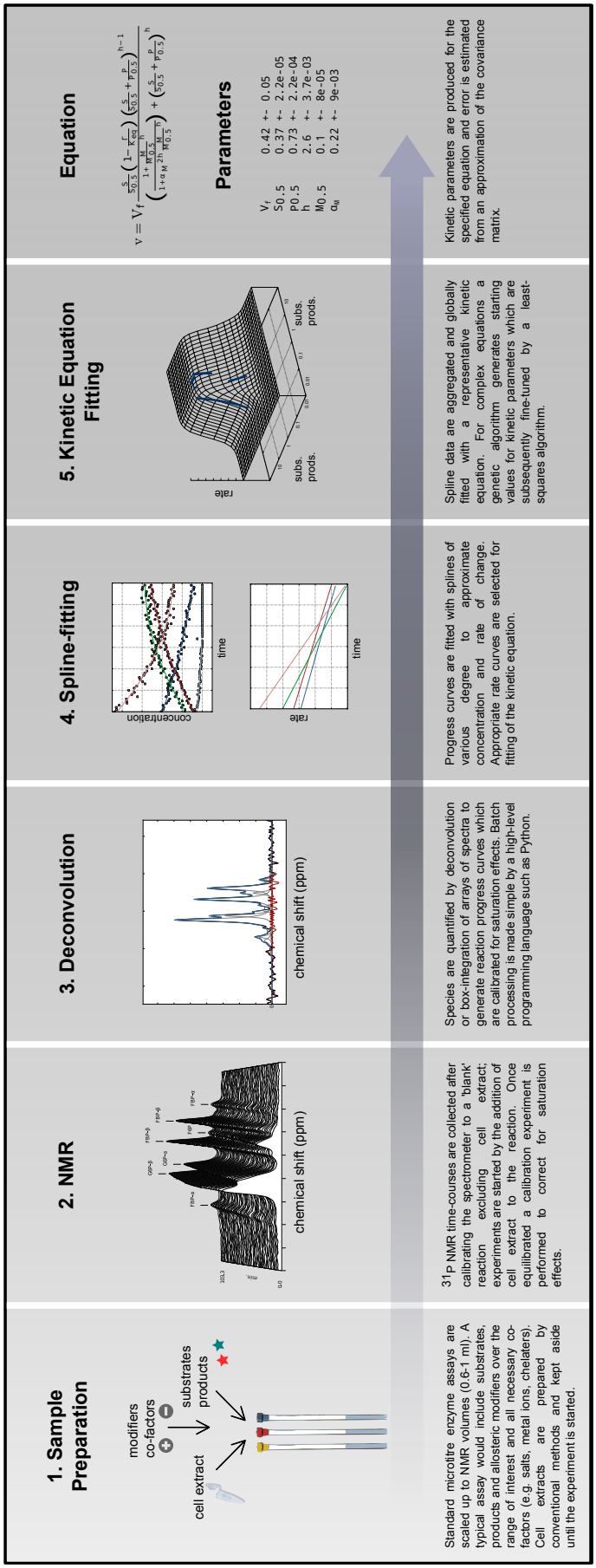
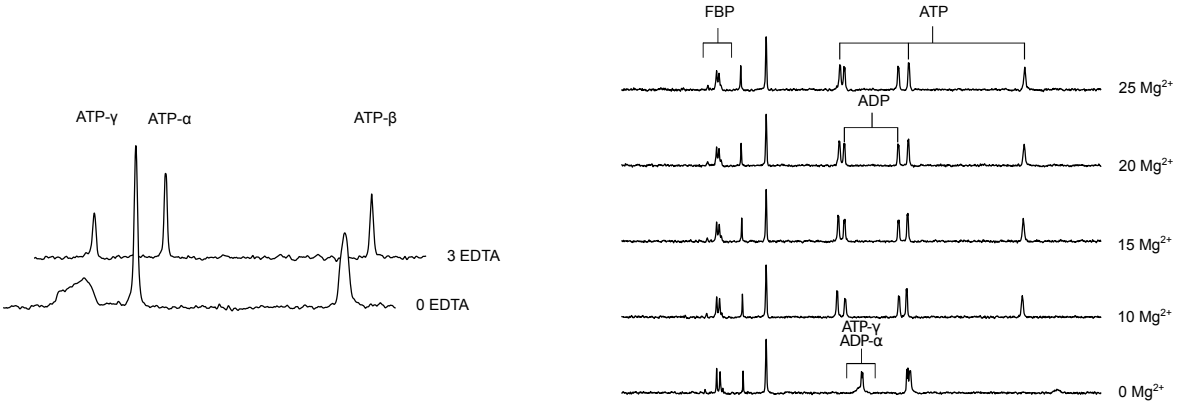


Figure 3.1: A work flow diagram.

obtained from various runs are then globally fitted with an appropriate kinetic equation to obtain kinetic parameters.

^{31}P NMR spectroscopy of nucleoside phosphates

Using ^{31}P-NMR to quantify ATP and ADP has to be approached carefully due to the diverse solvation, complexation and metal-binding behaviours of the nucleoside phosphates (see e.g. [191]). ATP binds Mg^{2+} to form the true substrate of PFK, MgATP [187]. Mg^{2+} concentration has a large effect on the NMR line shapes of ATP and ADP. Concentrations of Mg^{2+} up to and in the region of total nucleoside phosphate sharpen ATP and ADP line shapes, presumably by titrating out free nucleoside phosphate, but also shift the three resonances down field and reduce phosphorus coupling constants [192]. However, increasing cation concentration is offset against the concomitant loss in resolution in the rest of the NMR spectrum at high concentrations (Fig. 3.2b). Commercial preparations of ATP and ADP can contain trace amounts of metal ions, which significantly affect line shapes in a NMR spectrum and ideally should be removed by chelation before experimentation (e.g. using hydroxyquinoline or EDTA, Fig. 3.2a). In this study, prior to the addition of cell extract, PFK assay mixtures were treated with EDTA, after which excess Mg^{2+} was added to achieve the desired experimental concentration.



(a) (b)

Figure 3.2: ^{31}P NMR (a) The effect of EDTA on the line shapes of ATP using ^{31}P NMR. Spectra were collected with a 90° pulse angle and repetition time of 1 s (0.5 s acquisition time, 0.5 s relaxation delay). ATP concentration was 5 mM. (b) $MgCl_2$ titration of FBP, ATP and ADP and the effect on ^{31}P NMR spectral offset and line shape. Spectra were collected with a 60° pulse angle and repetition time of 1.3 s (0.8 s acquisition time, 0.5 s relaxation delay). FBP, ATP and ADP concentrations were 10 mM, and the indicated concentration of $MgCl_2$ (in mM) was added. All other parameters are described in Section 3.3.3. Raw NMR FID data are included as supplementary material.

Maximal rate normalisation

Due to the large number of variables involved in NMR spectrometry that are outside of human control (e.g. magnetic field inhomogeneities introduced by environmental changes such as the movement of metal items in the vicinity of the spectrometer) and the variable protein yield of cell extraction procedures (e.g. the high level of thermal denaturation introduced by sonication), it is often necessary to introduce an experimental normalisation factor to ensure that sets of assays performed on different extracts or days remain comparable. In this study, a maximal rate assay (i.e. saturated with substrate(s) and in the absence of products) was performed for each experimental session or independent extract, and all associated assays were normalised to this rate. Once aggregated and globally fitted, kinetic equations were scaled by an appropriate maximal rate acquired using either NMR or an independent spectrophotometric enzyme assay (in the case of PFK in this study both NMR and coupled enzyme assays were used to determine $V_{\max}$).

This approach is only valid, of course, if enzyme activities remain constant with time and there is no enzyme denaturation. As a control, we performed PFK assays at saturating substrate concentrations under conditions where these would not change significantly during the time course (Fig. 3.6, latter three datasets). Importantly, the rates were constant over the full time course (up to 60 minutes), demonstrating a lack of enzyme denaturation.

Data redundancy and model validation

Owing to the fact that multiple metabolites are simultaneously visible to the NMR spectrometer, NMR enzyme assays have a high data redundancy; progress curves for multiple enzymatic reactions can be generated from a single assay. In this study, progress curves used to fit the PFK reaction involved both a reverse PGI reaction and a forward PFK reaction—reactions were started with an initial concentration of F6P which was consumed in reverse by PGI and in the forward direction by PFK. PFK parameters were only fitted to data from the later stage of time courses, after PGI equilibration. This strategy provided two experimental benefits: first, lower concentrations of F6P, which are obscured by the adjacent accumulating FBP peaks, could be approximated assuming equilibrium with the easily-quantifiable G6P peaks; and second, the pre-equilibration time course data could be reserved for model validation by attempting to predict the behaviour of the coupled two-enzyme system using the independently fitted parameters of both PGI and PFK in a minimal model simulation (see Section 3.2.3).

3.2.2 Kinetic characterisation of phosphoglucose isomerase and phosphofructokinase

Phosphoglucose isomerase kinetic parameters

$$\text{G6P} \rightleftharpoons \text{F6P}$$

In order to characterise PGI, a total of five time courses in both forward and reverse directions of the reaction at different starting concentrations of substrate and product were collected, processed and fitted with splines to approximate concentrations and rates (Figs. 3.3a and 3.4). No "exit reactions" were observed and the reactions could be seen to proceed toward the literature equilibrium value ($K_{eq} = 0.28$ [193]). This was expected, as both enzymes adjacent to PGI, glucose 6-phosphate dehydrogenase and phosphofructokinase, require co-factors that were excluded from the assay mixture. Spline data were subsequently fitted with a uni-substrate/uni-product reversible Michaelis-Menten equation (Fig. 3.3b, for the equation see Table 3.1). This equation is the reduction of the Hill equation with $h = 1$, indicating a lack of cooperativity [83]. This approach was adopted as h consistently fitted with a value of ≈ 1 and is additionally justified by the absence of cooperativity in literature reports regarding PGI. Fitted kinetic parameters are summarised in Table 3.1.

Fitted half-saturation constants for G6P (0.550 ± 0.236 mM) and F6P (0.152 ± 0.017 mM), maximal forward rate (3.551 ± 0.050 µmol.min^{-1}.mg^{-1}) and the equilibrium constant ($0.286 \pm 8 \times 10^{-6}$) are comparable with literature values: $\text{G6P}_{0.5}$ 0.28 mM [195], $\text{F6P}_{0.5}$ 0.147 mM [195], V_f aerobic 3.29, micro-aerobic 4.66 µmol.min^{-1}.mg^{-1} [150], $K_{eq} = 0.28$ [193]. The higher margin of error on $\text{G6P}_{0.5}$ is most likely due to the sparsity of data at and below the half-saturating concentration of G6P (Fig. 3.4).

Phosphofructokinase kinetic parameters

$$\text{F6P} + \text{ATP} \rightleftharpoons \text{FBP} + \text{ADP}$$

^{31}P-NMR kinetic assays for PFK were performed primarily with no proton decoupling over a relatively wide spectral width of 10 to -25 ppm to include the nucleoside phosphates (Fig. 3.5). Six data sets were collected in total (Fig. 3.6). As a control for lack of fructose 1,6-bisphosphatase activity, NMR was performed on a cell extract incubated with FBP at 3 and 6 mM; no activity was observed (data not shown).

PFK NMR assays introduced a measure of complexity due to the following factors:

- Similarly to G6P, fructose 1,6-bisphosphate (FBP) exists as a pair of anomers in solution with the β-anomer predominating [196]. However, because of the two phosphate moieties, each anomer gives rise to two phosphorus peaks, and thus the molecule is observed as a quartet in the ^{31}P-NMR spectrum (Fig. 3.5b).

Table 3.1: Kinetic Equations and Fitted Parameters for the PGI and PFK reactions.

Kinetic Equations[a]		Fitted parameters[b]		
PGI	$v = V_{\mathrm{f}} \frac{\mathrm{g6p}\left(1-\frac{\Gamma}{K_{\mathrm{eq}}}\right)}{1+\mathrm{g6p}+\mathrm{f6p}}$	V_{f}	3.551 ± 0.050 ($V_r = 3.431$)[e]	µmol.min^{-1}.mg^{-1}
		$\mathrm{G6P}_{0.5}$	0.550 ± 0.236	mM
	uni-uni, reversible Michaelis-Menten[c] [83]	$\mathrm{F6P}_{0.5}$	0.152 ± 0.017	mM
		K_{eq}	0.286 ± 8×10^{-6}	
PFK	$v = V_{\mathrm{f}} \frac{\mathrm{f6p}^h \mathrm{atp}^h}{\left(\frac{1+\mathrm{pep}^h}{1+\alpha^{4h}\mathrm{pep}^h}\right)+\left(\frac{1+\alpha^{2h}\mathrm{pep}^h}{1+\alpha^{4h}\mathrm{pep}^h}\right)(\mathrm{f6p}^h+\mathrm{atp}^h)+(\mathrm{f6p}^h\mathrm{atp}^h)}$	V_{f}	0.4435 ±0.0001	µmol.min^{-1}.mg^{-1}
		$\mathrm{F6P}_{0.5}$	0.4174 ±0.0001	mM
		$\mathrm{ATP}_{0.5}$	0.5444 ±0.0003[d]	mM
	allosteric modifier: ⊖PEP	$\mathrm{PEP}_{0.5}$	0.0863 ±0.0001	mM
	bi-substrate, irreversible Hill [88]	α	0.3797 ±0.0001	
		h	1.8830 ±0.0019	

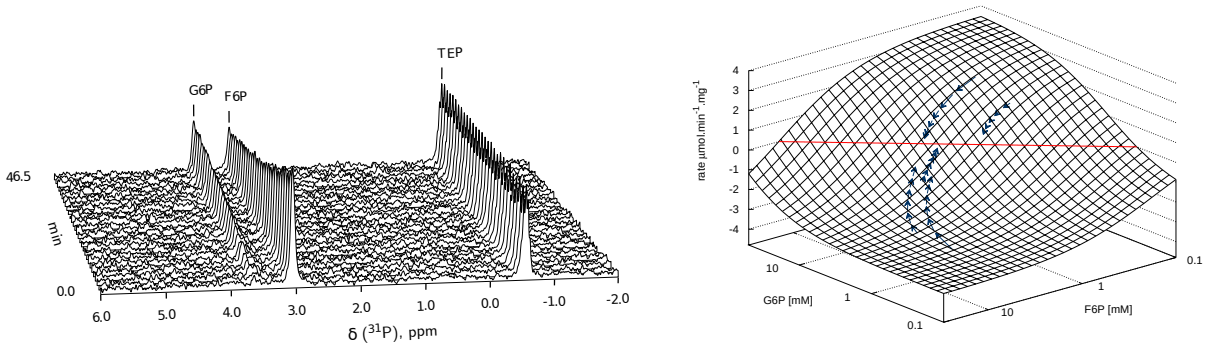
[a] Metabolites are scaled by their half-saturation constants: e.g. $\mathrm{g6p} = \frac{[\mathrm{G6P}]}{\mathrm{G6P}_{0.5}}$. Γ is the mass action ratio of unscaled concentrations: $\frac{[\mathrm{F6P}]}{[\mathrm{G6P}]}$ for PGI, $\frac{[\mathrm{FBP}][\mathrm{ADP}]}{[\mathrm{F6P}][\mathrm{ATP}]}$ for PFK.

[b] Error is Standard Error of the Mean (S.E.M.): variance in parameter $\times \frac{\text{sum−squared error}}{\text{degrees of freedom}}$. A covariance matrix is derived from a Jacobian (a matrix of first-order partial derivatives) approximation to the Hessian matrix (a matrix of second-order partial derivatives describing the curvature of the objective function) around the solution and the associated variance in the fitted parameters is scaled by the residual variance. This correction of the estimated variance produces an unbiased estimator of the spread of the fitted parameters by scaling the residuals so that they are in units of standard deviations as described in the `scipy.optimize.leastsq` documentation [163]. For a fuller treatment of the Levenberg-Marquardt algorithm see [194].

[c] In the case of no cooperativity ($h = 1$), the reversible Hill equation reduces to the reversible Michaelis-Menten.

[d] Correcting for low Mg^{2+} value as used in this study (see Results and Discussion) gives $0.1089 \pm 1 \times 10^{-5}$

[e] Inferred from the Haldane relationship [83].



(a) (b)

Figure 3.3: Phosphoglucose Isomerase. (a) Example of a ^{31}P-NMR time course of a PGI reaction using a cell extract incubated with an initial concentration of 8.5 mM F6P and no G6P, collected at a 60° pulse angle over 47 min (0.8 s acquisition, 0.5 s relaxation). Additional NMR parameters are described in Section 3.3.3. In this reaction F6P was converted in reverse to G6P as the reaction approached equilibrium. The time course is not shown to full equilibration; final concentrations were 5 and 2.8 mM for G6P and F6P respectively. TEP is an internal standard. (b) Reversible Michaelis-Menten equation (see Table 3.1) fitted to PGI progress curves derived from NMR data: equilibrium values are represented by the red contour line (—), arrows indicate both the metabolite concentrations and the direction of reaction as each time course progresses towards equilibrium (→→→). The rate was normalised to total protein concentration. Substrate and product concentration axes are in logarithmic scale. $R^2 = 0.99$.

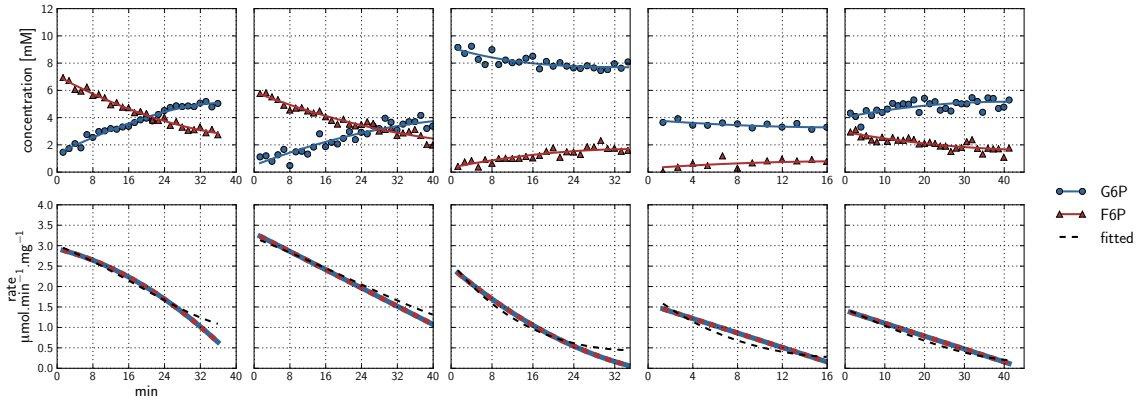


Figure 3.4: Spline fits of Phosphoglucose Isomerase data. **Top row:** Time courses of the PGI reaction were acquired by incubating a cell extract with various starting concentrations of substrate G6P (■) and product F6P (▲) and monitoring reaction progress using ^{31}P NMR with a 90° pulse angle and 1 s repetition time (1.0 s acquisition, 0.0 s relaxation) with 80 transients per FID. Other parameters are as described in Section 3.3.3. Progress curves derived from NMR peak integrals were fitted with splines (G6P —, F6P —). **Bottom row:** The respective averaged absolute rates of the fitted splines are plotted (dual colours indicate average of two respective rates) with the rate of the fitted kinetic equation included (- - -, Table 3.1: PGI). Rates were normalised to total protein content. Raw NMR FID data, as well as NMR peak integrals and spline data, are included as supplementary material.

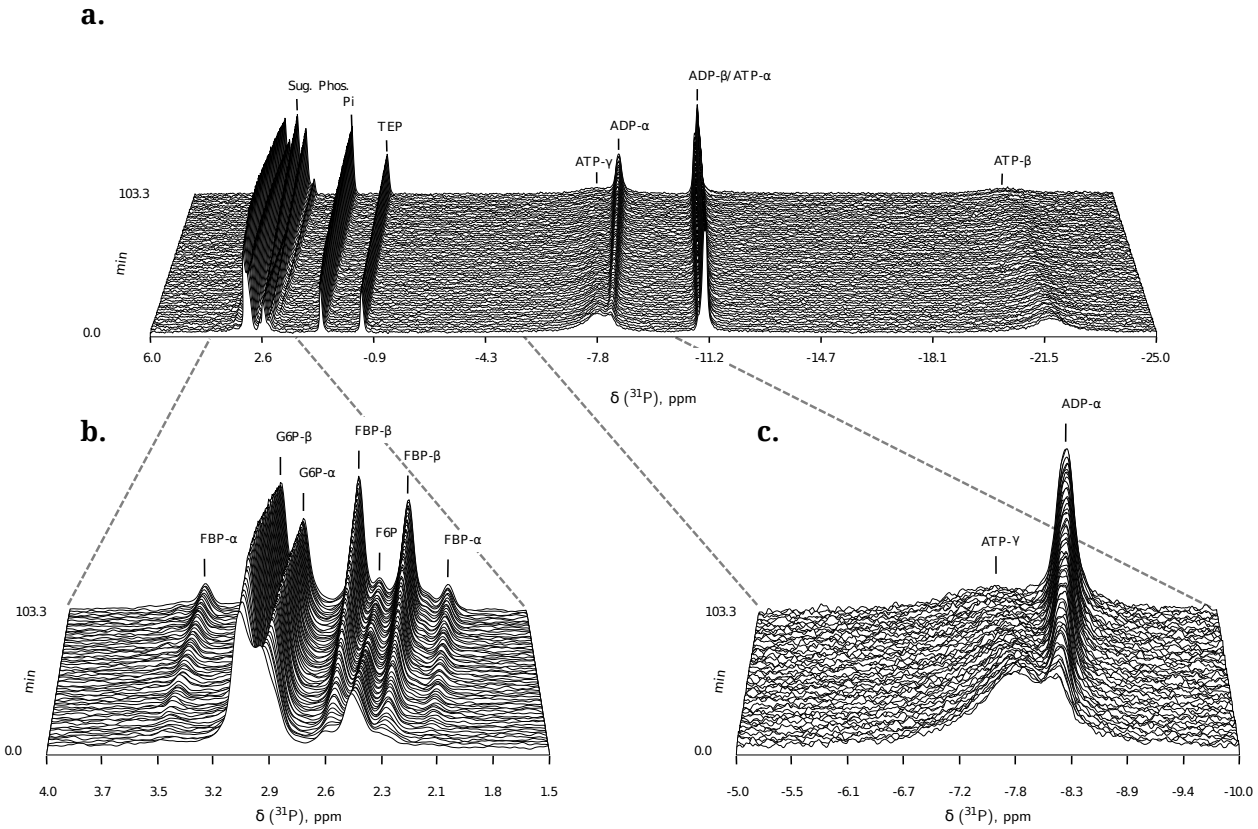


Figure 3.5: Example of a Phosphofructokinase ^{31}P-NMR time course. As high Mg^{2+} concentrations can lead to line-broadening and an obscured spectrum, data were collected with no additional Mg^{2+} (beyond the trace amounts left from the growth medium) for better resolution. A pulse angle of 60° and a repetition time of 1.3 s (0.8 s acquisition time, 0.5 s relaxation delay) was used. 10 mM triethyl phosphate (TEP) is included as an internal standard. All other parameters are described in Section 3.3.3. **a.** Full NMR spectrum. Initial concentrations were 14 mM G6P, 3 mM F6P, 13 mM ATP. The first few FIDs collected before the lock signal had stabilised, have been excluded. **b.** Expansion of the sugar-phosphate region (4.0 to 1.5 ppm). **c.** Expansion of the nucleoside phosphate region (−5 to −10 ppm).

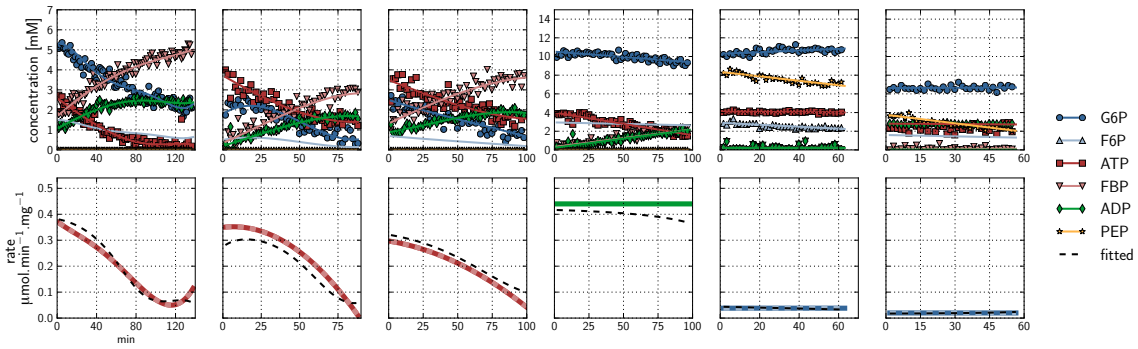


Figure 3.6: Spline fits of ^{31}P-NMR Phosphofructokinase data. Data were acquired using a 90° pulse angle to collect 100 transients per FID using a repetition time of 1 s (0.5 s acquisition time, 0.5 s relaxation delay). **Top row:** Progress curves representing NMR peak integrals (G6P ●, ATP ▲, FBP ▼, ADP ◆, PEP ★) are fitted with splines (G6P —, F6P —, ATP —, FBP —, ADP —, PEP —). Inhibitor assays containing PEP are shown in the last two blocks. Note that with the exception of the second-last assay, F6P concentrations are inferred from equilibrium with G6P via PGI. **Bottom row**: Respective rates derived from spline-fitted NMR data. Dual colour lines indicate an average of two respective rates. For comparison, the rate calculated by the irreversible Hill equation (- - -, Table 3.1: PFK) at the specific substrate, product and effector concentrations is shown. The Hill equation parameters were the same throughout and obtained from a global fit of all the time courses shown. Rates are normalised to total protein. Raw NMR FID data, as well as NMR peak integrals and spline data, are included as supplementary material.

- An additional complexity is that F6P appears between the two peaks (2.6, 2.3 ppm) of the FBP β-anomer at ~2.4 ppm. At low F6P and high FBP concentrations, typical of late-stage PFK time courses, F6P is obscured by the FBP peaks and has to be estimated by assuming equilibrium with the easily quantifiable G6P via the much faster PGI reaction. This is a reasonable approximation, provided that the PGI reaction is allowed to equilibrate before data acquisition (the maximal rate of PGI is ~7.5 times that of PFK). In all experiments, PGI was active, and thus to maintain higher concentrations of F6P, at times near-equilibrium concentrations of G6P were added. Data collected before PGI equilibration were excluded from fitting, reserving them for validation (Section 3.2.3).

- FBP-aldolase activity was not observed. This was to be expected as aldolase from *E. coli* is strictly Zn^{2+}-dependent [197] and Zn^{2+} was excluded from assay mixtures.

PEP was included in two assays at concentrations of 8 and 4 mM to assay for inhibition (Fig. 3.6). Metabolism of PEP to 2-phosphoglycerate and 3-phosphoglycerate via the enolase and phosphoglycerate mutase reactions could be observed. As the PFK reaction progressed, ADP was produced, which was subsequently consumed by pyruvate kinase, providing another exit route for PEP. The result was a significant decline in the PEP concentration over the duration of the experiment, and a concomitant maintenance of the ATP concentration (Fig. 3.6). This dynamic concentration change is fortuitous for fitting purposes, as it can eliminate the need for performing several assays at different static effector concentrations.

As a kinase and "committed step" of glycolysis, the forward reaction of PFK is favoured (K_{eq} = 800) [198]. In a time course assay where the reaction is allowed to equilibrate, it becomes impossible to collect data for the reverse reaction when the K_{eq} is too high, as such concentrations fall below the detection limits of NMR. PFK was therefore fitted with an irreversible bi-substrate Hill equation with PEP as a negative allosteric modifier (Fig. 3.7, see Table 3.1 for the kinetic equation and all fitted parameters). As per standard practice for these experiments, time courses were normalised by a maximal rate for a given experimental day or cell extract, to reduce the possible introduction of error between these marginally different experimental conditions (see Section 3.2.1). The $V_{\max}$ for PFK was thus subsequently determined using a coupled enzyme assay (normalised to total protein content) and was identical to the highest NMR-determined rates.

The maximal forward rate was determined using a spectrophotometric assay in which PFK activity was linked to NADH consumption by α-glycerophosphate dehydrogenase via fructose-1,6-bisphosphate aldolase and triosephosphate isomerase (Section 3.3.5). The maximal rate was determined as 0.4435 ± 0.0001 $\mu mol.min^{-1}.mg^{-1}$, assuming negligible activity of glyceraldehyde 3-phosphate dehydrogenase. This rate is very similar to the maximal rates determined by NMR and falls within literature range: 0.34 (aerobic) – 0.54 (micro-aerobic) $\mu mol.min^{-1}.mg^{-1}$ [150, 187].

In the absence of significant nucleoside diphosphate, the binding of F6P and ATP to PFK is that of a bi-reactant random sequential enzyme in rapid equilibrium, which displays significant antagonistic binding between the substrates [189]. Under these conditions, the binding of F6P is

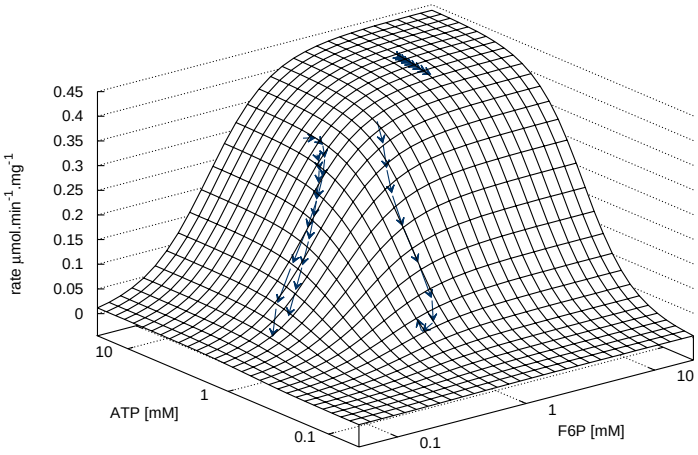


Figure 3.7: Phosphofructokinase: irreversible bi-substrate Hill equation globally-fitted to aggregated ^{31}P-NMR progress curves (see Table 3.1 for equation and fitted parameters). Rate is normalised to total protein. Substrate concentration axes are in logarithmic scale. Arrows indicate both the metabolite concentrations and the direction of the reaction for individual time courses (→→→). PEP inhibitor assay data have been excluded from the plot as only two variables can be visualised simultaneously. $R^2 = 0.96$.

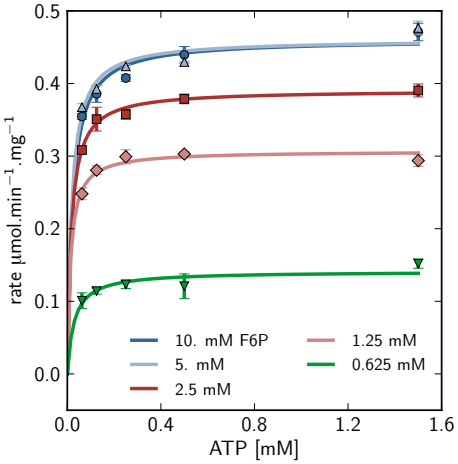
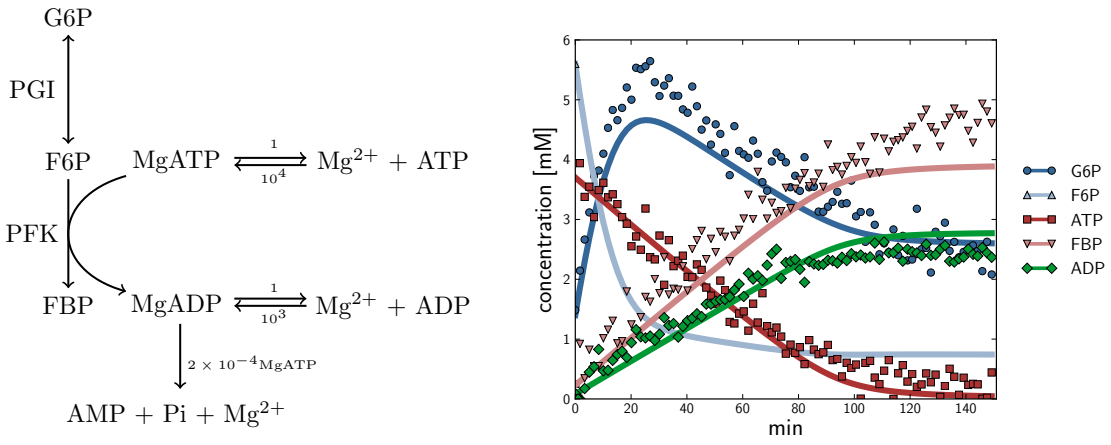


Figure 3.8: Phosphofructokinase enzyme-coupled kinetic assay. ATP saturation curves at different F6P concentrations were generated using the coupled enzyme assay system as described in Section 3.3.5. Points represent initial rate data and are fitted with a standard irreversible Michaelis-Menten kinetic equation. Error bars represent experimental replicates ($n = 3$) and are S.E.M. F6P: 10 mM (-●-), 5 mM (-▲-), 2.5 mM (-■-), 1.25 mM (-◆-), 0.625 mM (-▼-).



(a) (b)

Figure 3.9: A simulation of a two-enzyme NMR time course involving PGI and PFK, beginning with initial measured concentrations. (a) Model schematic. Kinetic parameters are as described in the text. To accurately approximate experimental conditions, the model consisted of three reactions in addition to the two glycolytic enzymes: both ATP and ADP were in rapid-equilibrium reactions with MgATP and MgADP (1 mM free Mg^{2+}, K_{eq} values were 10^4 and 10^3 respectively [199, 200]), and MgADP was consumed by an elementary first-order hydrolysis reaction producing AMP + Pi + Mg^{2+} ($k = 2 \times 10^{-4}$); see text for details. (b) Simulated time course concentrations (G6P —, F6P —, ATP —, FBP —, ADP —) compared to experimental time course data (G6P ●, F6P ▲, ATP ■, FBP ▼, ADP ◆). The time course started with only F6P and ATP present as substrates. Note: for parameter fitting F6P was assumed to be in equilibrium with G6P *via* the PGI reaction, and as such no quantified F6P data are included in this figure, except for an initial concentration. AMP and orthophosphate are not shown.

cooperative with a half-saturation constant of 0.35 mM at pH 8.5 [187, 189]. This is very close to the fitted parameter of ≈ 0.42 mM. The fitted Hill coefficient of ≈ 1.9 indicates significant cooperativity and is, as expected, lower than literature values due to the accumulation of ADP during the time course, which abolishes cooperativity; product was not allowed to accumulate in the cited studies [187, 189]. ATP-binding is unaffected by the presence or absence of nucleoside diphosphates. The half-saturation concentration for ATP, however, has been shown to change over the range 0.01–0.16 mM from low to high F6P concentrations, respectively [189]. The fitted $ATP_{0.5}$ value of ≈ 0.54 mM was significantly higher than these values. For independent verification using an alternate method, a series of coupled enzyme assays was performed by varying ATP concentration over the range 0.0625–1.5 mM at five different F6P concentrations: 0.625, 1.25 , 2.5 , 5 and 10 mM (Fig. 3.8). The resulting ATP half-saturation concentrations can be seen to be lower than the NMR fitted parameter values, in agreement with the literature values.

A possible explanation for the fact that the fitted NMR parameter for $ATP_{0.5}$ was higher than the value obtained in the coupled enzyme assay, is that the Mg^{2+} concentration was kept lower

than the total ATP concentration (1 mM Mg^{2+} *vs.* 2.5–5 mM ATP) to retain resolution in the NMR spectra (see Section 3.2.1). Since the true substrate of PFK is MgATP, the effective substrate concentration was therefore significantly lower than the added ATP. Re-fitting the data assuming a five-fold lower concentration of ATP (due to saturation of available Mg^{2+} and competition for Mg^{2+} by increasing ADP concentrations) reduced the $ATP_{0.5}$ value to 0.1089 ± 0.0001, which is well within the literature range (other parameter values showed no significant change).

In a PFK assay system using purified enzyme with no additional reactions one would expect products for the bi-substrate reaction (FBP and ADP) to accumulate at identical rates. However, using a whole cell extract as in the current method, unidentified background reactions could consume one or both of the products. This phenomenon was observed as a slower accumulation of ADP compared to FBP (Fig. 3.6, first three time courses). The following proposed pathway could explain this effect:

1. PFK	$\mathrm{ATP} + \mathrm{F6P} \rightarrow \mathrm{FBP} + \mathrm{ADP}$
2. ATP hydrolysis	$x\mathrm{ATP} \rightarrow x\mathrm{ADP} + x\mathrm{P_i}$
3. Adenylate Kinase	$2x\mathrm{ADP} \rightleftharpoons x\mathrm{ATP} + x\mathrm{AMP}$
Net Reaction:	$\mathrm{ATP} + \mathrm{F6P} \rightarrow \mathrm{FBP} + (1-x)\mathrm{ADP} + x\mathrm{AMP} + x\mathrm{P_i}$

In the scheme above, background ATP hydrolysis is compensated for by the house-keeping adenylate kinase reaction, which maintains adenylates in equilibrium. Adenylate kinase will proceed in the forward direction as shown due to the initial virtually absent AMP concentration. The net reaction was observed in NMR time courses as an increase in AMP and phosphate at concentrations visibly similar to the difference between FBP and ADP. No net ADP consumption was observed in assays containing PEP, as pyruvate kinase scavenged the available ADP and phosphorylated it to produce ATP (Fig. 3.6, latter two time courses).

3.2.3 Method validation: a minimal model of coupled reactions

To explore the validity of the experimental method presented here and the accuracy of the fitted kinetic parameters, the parameterised equations were evaluated by the construction of a minimal model of the 2-enzyme system under study.

A model representing a NMR time course of the PGI and PFK reactions, taking MgATP and MgADP association reactions into account, was constructed (Fig. 3.9a) and simulated over time from measured initial concentrations (Fig. 3.9b; time course experimental data are the same as in Fig. 3.6, top left), using the PySCeS (Python Simulator for Cellular Systems) software [154] . The earlier portion of the time course (0–30 min, before F6P and G6P were equilibrated via the PGI reaction), which was excluded from parameter fitting, was included in this instance (see Section 3.2.1). The reaction begins with F6P and ATP; F6P is consumed in the reverse PGI reaction to produce G6P until equilibrium is reached; F6P and ATP are consumed by the forward

PFK reaction. To mimic the net reaction mentioned in Section 3.2.2 and account for non-specific ATP hydrolysis, MgADP consumption was included as a first order reaction dependent on MgATP ($k = 2 \times 10^{-4}$, not fitted). The fitted kinetic parameters are able to predict correctly the changes in metabolite levels for this two-enzyme system, demonstrating the adequacy of the presented methodology for systems biology applications.

3.2.4 Comparison with other approaches

A key methodological goal of systems biology is the development of techniques to assay enzyme kinetics *in vivo*. As a non-invasive technique, NMR technology is being developed with this application in mind and *in vivo/in situ*[1] fluxes are becoming quantifiable [19]. Unlike traditional methods, NMR spectroscopy allows for the simultaneous observation of many different metabolites in an enzyme assay, generating a multiplicity of reaction rates determined only by the experimental starting conditions. There are several experimental benefits to this design:

1. A single assay can be designed to produce rate and substrate concentration data for multiple enzymatic reactions, reducing time, cost and labour. In this study, a number of the datasets used in parameter fitting of the PFK reaction are time courses of both the PGI and PFK reactions.

2. Provided a NMR-sensitive nucleus is present (^{31}P in this instance), all substrates, products and effectors can be quantified in real time. This simultaneous quantification of all metabolites circumvents an important caveat of traditional enzyme kinetics. Often metabolites and effectors will be consumed or produced by ancillary reactions mediated by enzymes other than those being studied (or simply uncatalysed reactions), a phenomenon mostly invisible to traditional enzyme assay techniques. In this study, the PFK datasets exhibited this phenomenon: F6P was consumed in reverse by the preceding glycolytic enzyme PGI producing G6P; ADP, which is both a product of the PFK reaction and exhibits a complex allosteric relationship to the PFK enzyme, was consumed by a proposed hydrolytic reaction scheme, producing AMP and orthophosphate; the allosteric inhibitor PEP was consumed both in reverse by the enolase and phosphoglycerate mutase reactions producing 2-phosphoglycerate and 3-phosphoglycerate, and in the forward direction by the pyruvate kinase reaction as ADP was released from the PFK reaction, maintaining ATP levels and generating pyruvate. Though these ancillary reactions are also taking place in the NMR time course assays, they are observable and can be taken into account during the data analysis.

3. When the concentration of an allosteric modifier changes during the experiment, this reduces the amount of data needed to fit allosteric kinetic equations by essentially providing an innate perturbation of effector concentration. In comparison, initial-rate enzyme assays require many

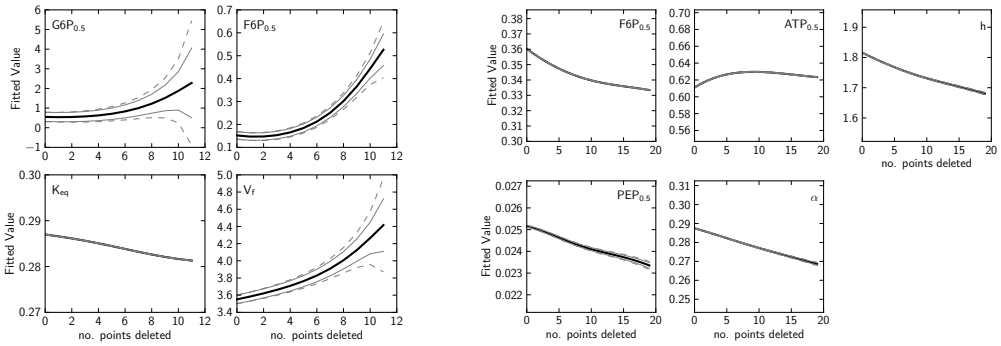
[1] *in situ* here refers to permeabilised cells.

reactions over a range of effector concentrations to achieve the same result; a difficulty that is exponentially compounded by the presence of multiple effectors.

Progress curve fitting using an integrated Michaelis-Menten equation (e.g. the Lambert-W form [176]) exhibits a number of benefits over the traditional methods (see Introduction), not least of which is the fact that the technique utilises the full time course dataset in which substrates vary in a dependent fashion, rather than merely utilising the initial rates of reactions (a notoriously difficult portion of the progress curve to measure accurately [179]). However, this methodology does suffer from a number of drawbacks. Stated simply, not all of the possible causes of a change in reaction rate are distinguishable from an enzymatic time course [179]. Product inhibition is difficult to account for in progress curve analyses and is circumvented only by combining the results of numerous assays or by using a technique such as NMR, which provides real-time analysis of metabolites including reaction products. This difficulty is compounded by the presence of multiple products as well as the lack of quantifiability of products and effectors discussed above, as the shape of the progress curve is dependent not only on gradual changes in substrate concentration, but can also be altered by changing product and effector concentrations as they are consumed or produced by invisible side reactions [174]. To assess whether a particular enzyme is suitable for progress curve analysis, a simple assay has been suggested that indicates the presence of side reactions, substrate/product inactivation and enzyme instability [201].

The technique of NMR spectroscopy overcomes many of these handicaps. However, it must be stressed that enzyme instability should be considered when attempting progress-curve analysis. For example, reactions should produce linear rates under saturating conditions (without significant product accumulation) over the timespan of the experiment. Also, as vital kinetic information is extracted from progress curves around half-saturation concentrations and near equilibrium, NMR spectroscopy may not provide sufficient sensitivity to estimate all kinetic parameters associated with reactions catalysed by enzymes with extremely low half-saturation constants (< 0.5 mM) or extreme equilibrium constants.

To investigate the effect of varying the duration of the time course during progress curve analysis, and to assess the requirement for near-equilibrium data, PGI and PFK spline-fitted data were sequentially truncated at the end of the time course and refitted with the standard fitting routine (Fig. 3.10). The total number of truncated data points represents a portion of the closer-to-equilibrium side of each time course in the two data sets. These truncations represent up to 40 % (Fig. 3.10a) and 24 % (Fig. 3.10b) of the longest time course for PGI and PFK, respectively. Fitted parameters remained very stable after deleting at least the first half of the truncated data. It is also clear from the estimated error that as the time courses proceed towards equilibrium, the fits converge upon the parameters previously fitted and remain stable for roughly the first half of truncation. This demonstrates that the data were collected over sufficiently long time periods to include the necessary information for parameter estimation (both in terms of changing substrate and product concentrations, and in terms of thermodynamic detail on the approach to equilibrium



(a) (b)

Figure 3.10: (a) Phosphoglucose Isomerase and (b) Phosphofructokinase parameter fitting performed after deleting a number of closer-to-equilibrium data points. To observe the effect of losing data from the latter part of the time courses on the fitting process, parameter fitting was performed as before using the Levenberg-Marquardt algorithm after a series of truncations had been made to the spline-fitted datasets, starting at the end of each time course, i.e. closer to equilibrium (original data in Figs. 3.4 and 3.6; fitted kinetic equations as in Table 3.1). Error was calculated as before (Table 3.1, footnote b) but with two variations: the first method involved rescaling the degrees of freedom to reflect the degree of truncation of the data (- -), the second method retained the original degrees of freedom (—). This approach to error estimation was adopted to be able to distinguish between two sources of error: that due to losing data generally, and that due to specifically losing closer-to-equilibrium data. Twelve and twenty data points were deleted sequentially from the PGI and PFK data, respectively (representing up to 40 % and 24 % of the length of the longest respective time courses). Fitted parameters (—).

in the case of PGI). Truncation of the data and the subsequent explosion of the error margins on most of the PGI parameters (and the comparative stability of the PFK parameters) suggests that near-equilibrium data is indeed essential when fitting reversible kinetic equations (but of course not used for fitting irreversible equations).

In summary, we have shown that globally fitting a collection of progress curves generated with NMR spectroscopy—a technique that allows comprehensive and simultaneous quantification of metabolites—with a generic rate equation is able to produce adequate enzyme kinetic parameters for modelling biological systems. Moreover the method presented here overcomes most of the difficulties presented by traditional enzyme assays (including progress curve analyses), while being materially inexpensive, less labour-intensive and relatively rapid.

3.3 Experimental section

3.3.1 Growth conditions and Media

Escherichia coli K12 W3110 was used to create cell stocks for kinetic assays by growing a 1-l batch culture inoculated to an optical density at 600 nm (OD_{600}) of 0.1 with an overnight starter culture in M9 minimal medium [0.4% glucose, 1.28% Na_2HPO_4 (w/v), 0.30% KH_2PO_4 (w/v), 0.05% NaCl (w/v), 0.10% NH_4Cl (w/v), 0.05% $MgSO_4$ (w/v), 0.001% $CaCl_2$] at 37 °C and pH 7.2, for the strict physiological control required for enzyme assays (typical constituents of rich media, such as yeast extract or tryptone digest, are highly complex and may exert any number of influences on the metabolic network). Cells were grown at pH 7.2 to buffer against the production of acidic fermentation products and to prevent premature pH-inhibition of glycolysis [202]. Cultures were gently mixed with magnetic stirrer bars during growth to produce a micro-aerobic environment. Cells were harvested in mid-log phase (OD_{600} = 0.45) and centrifuged for 10 min at 5,000 rpm (∼4200 × g). Pellets were combined and resuspended in 100 mM PIPES buffer (pH 7.2) to 50 ml as a washing step and centrifuged at 5,000 rpm (∼5,600 × g) for 10 min. This pellet was resuspended in 20 ml of 100 mM PIPES (pH 7.2) and separated into twenty 1-ml aliquots in Eppendorf tubes. After microcentrifugation for 10 min at 13,000 rpm (∼ 10,000 × g), the supernatant was discarded and cell pellets were frozen in liquid N_2. These stocks were maintained at −80°C. Pellets were kept on ice between harvesting steps.

3.3.2 Extraction

To produce a whole-cell extract, frozen cell stocks were thawed and resuspended in 1 ml 100 mM PIPES buffer (pH 7.2) with 1 mM phenylmethanesulphonylfluoride (PMSF) to inhibit serine-protease activity. Cells were extracted either by sonication or with a glass-bead shaking method (a gentler method better suited to accurate enzyme activity measurements). Protein concentration was determined by the Bradford assay method [203]. Both methods were used in this study, however glass-bead extracts were used for velocity determination. Sonication is a faster method but suffers the drawback of thermal denaturation, which can reduce absolute maximal rate achieved by the extracted enzymes; though this is unlikely to affect kinetic parameters such as half-saturation constants.

Sonication

Resuspended cell pellets in a 2 ml Eppendorf tube were placed in an ice slurry and sonicated at 30 s intervals with 15 s breaks (to prevent overheating) for a total sonication time of 4 min using a micro-tip at 138 kPa. To determine the optimal sonication time that maximises protein yield without denaturing the enzymes of interest, a series of extractions was performed for various times and the activity of lactate dehydrogenase (LDH) was assayed. Optimal specific activity (μmol.min^{-1}.mg^{-1}) was achieved at 4 min sonication time.

Glass-bead extraction

1.5 g $\leq$ 106 μm diameter glass beads (Sigma) were added to a 1 ml resuspended cell pellet in a 50 ml Falcon tube and rotated on an orbital shaker at 450 rpm for 1 h. This method has been shown to retain higher enzyme activity levels when compared with other methods and thus was used for maximal velocity measurements [204]. Protein and activity yield was optimised using the LDH assay.

3.3.3 NMR spectroscopy

All reaction components were from Sigma-Aldrich (except ATP from Boehringer-Mannheim) and prepared in 100 mM PIPES (pH 7.2, corrected by the addition of 10 M NaOH). Triethyl phosphate (TEP) was introduced as an internal standard due to its metabolic inertness [205]. A standard assay in a 5 mm glass tube was composed of: 50 mM TEP, 100 μl D_2O, 100 μl cell extract, 100 μl per substrate/product/co-factor, 1 mM Mg^{2+}, and filled to a final volume of 1 ml with 100 mM PIPES (pH 7.2). Initially, cell extract was excluded. This blank was used to tune the spectrometer, acquire a lock signal and shim the instrument before the reaction was started by removing the tube from the instrument, adding 100 μl cell extract, mixing several times by inverting, and re-inserting the tube. Data acquisition was initiated once a stable lock signal was achieved.

^{31}P NMR was performed at 25 °C and a frequency of 242.87 MHz on a Varian 600 MHz spectrometer with a 1 s repetition time (1.0 s acquisition/0.0 s relaxation for PGI, 0.5 s acquisition/0.5 s relaxation for PFK) to collect 80 (PGI) or 100 (PFK) transients per FID using a pulse angle of 90° with either no proton decoupling or a low power decoupling (Waltz-16) to prevent overheating of the sample. In rapid-sampling NMR it is typically not possible to accommodate the full spin-lattice relaxation of the nucleus of interest. T1 relaxation times were determined for the metabolites in this study and varied from 0.2 s (nucleoside phosphates) to 6 s (TEP). Species concentrations were thus calibrated for incomplete relaxation with a fully-relaxed spectrum (30 s relaxation time, 5 $\times$ the longest T1) of a cocktail of metabolites of interest.

Time courses collected using a particular cell extract or on a certain day were normalised by a representative maximal rate at saturating substrate concentrations from that NMR session. This was done in order to prevent the introduction of error between days and extracts.

Where ATP and ADP were unquantifiable due to metal ion contamination, 1.5 mM EDTA and 2.5 mM $MgCl_2$ were added sequentially to a final effective concentration of 1 mM Mg^{2+}.

3.3.4 Data processing

All data processing was performed using a custom software module[2] written in the Python programming language (the Varian data import function was taken from [161]; released under the Open Software Initiative and the New BSD Licence). Spectra were processed with an exponential

[2]Software available from authors on request.

line-broadening function of 8.5 Hz and peaks were quantified either by box-integration or deconvolution by fitting with Lorentzian functions. Quadratic splines were fitted to concentration data using the `scipy.interpolate` module [163]. All fitting procedures were performed using either a custom genetic algorithm or the Levenberg-Marquardt least-squares algorithm as employed in the `scipy.optimise.leastsq` module [163]. Images were produced using Matplotlib [162], Gnuplot [206] and Inkscape [207].

3.3.5 Enzyme assays

A coupled enzyme assay system was used for PFK with the following composition: 0.625–10 mM F6P, 0.062–1.5 mM ATP, 0.2 mM NADH, 10 mM $MgCl_2$, 1.5 U.ml^{-1} FBP-aldolase, 5 U.ml^{-1} triose-phosphate isomerase, 4.3 U.ml^{-1} α-glycerophosphate dehydrogenase. NADH consumption by α-glycerophosphate dehydrogenase was monitored at a wavelength of 340 nm in a 96-well plate reader (VarioSkan Microplate Reader, Thermo Electron Corp.; Greiner Bio-one Flat-bottom microplate) with the temperature maintained at 25°C. Activities were normalised with protein content determined by the Bradford assay [203]. All assays were performed in triplicate in 100 mM PIPES buffer at pH 7.2.

3.4 Supplementary data

The following data files are available at the *Metabolites* website (`http://www.mdpi.com/2218-1989/2/4/818/s1`):

pgi_fids.zip ZIP-file containing the raw Varian NMR FID data for all the PGI runs, plus a text listing of the various experiments.

pfk_fids.zip ZIP-file containing the raw Varian NMR FID data for all the PFK runs, plus a text listing of the various experiments.

supp_data.ods ODS-spreadsheet containing the integrated NMR peak data for all the experiments, as well as the fitted spline data for the concentrations and rates.

4 Construction and validation of a microaerobic model of glycolysis in *E. coli*

Summary

This chapter describes the construction and validation of a mathematical model of glycolysis in fermenting *E. coli* from four submodels: 1. a published model of the PTS sugar import pathway, 2. PGI and PFK as modelled in Chapter 3, with FBP aldolase, 3. a model for the lower half of glycolysis generated using *in vitro* NMR reaction time courses to approximate enzyme kinetic parameters *in vivo*, and 4. a constructed model of the fermentation pathways parameterised with published parameter values. Cells were grown and assayed under microaerobic conditions representing a more natural environment for the bacterium (e.g. in the lumen of the human small intestine) than, for instance, the artificial aerobic or anaerobic chemostat, and an environment inevitably encountered during batch culture as accumulating CO_2 produced during respiration gradually insulates the growing culture from exposure to O_2. To generate parameters for the bottom half of glycolysis (triosephosphate isomerase $\rightarrow$ pyruvate kinase) leading to pyruvate, a selection of submodels reflecting experimental conditions was fitted to *in vitro* ^{31}P NMR reaction time series. The resulting models went through a model selection process. Selected putative models were evaluated by comparing their simulated behaviour to *in situ* ^{13}C NMR time series using permeabilised cells under experimentally similar conditions. A final selection of three models was made, and these were expanded using literature values where available to represent the full spectrum of reactions of fermenting central carbon metabolism, including the glucose importing phosphotransferase pathway (PTS), and the fermentation reactions including the acetate, ethanol, lactate and succinate pathways. ^{13}C NMR assays of quenched fermentations were used to determine fermentative fluxes *in vitro*, and *in vivo* ^{31}P NMR was used to observe intracellular metabolite accumulation, and steady-state model behaviour was compared with published steady-state metabolite concentrations and fluxes.

4.1 Glycolysis: Regulation of a Micro-economy

Glycolysis produces ATP to meet the cell's needs for a diverse range of processes including growth, motility and reproduction. Economics provides a useful paradigm for understanding metabolic pathways as supply and demand systems balanced around an equilibrium, poised to respond to changing environmental conditions [99]. Glycolytic flux (as measured by glucose uptake) in aerobically grown *E. coli* has been shown to be controlled by the demand for ATP [101], that is, reactions external to glycolysis which affect the energy charge (or $\frac{[ATP]}{[ADP]}$ ratio) exert $> 75\%$ of the control over glycolytic flux.

Yet, thorough as that study is, it represents one of many studies broaching a topic of considerable complexity, of which a number are mentioned here. It has been argued that there is an inverse correlation between respiratory activity and uptake by the PTS: under aerobic conditions, substrates that stimulate respiration result in the inhibition of PTS activity [208–212]. Additionally, the inhibition of uptake by the PTS (usually assayed with the non-metabolisable glucose analogue α-methyl glucoside, α-MG) is abolished by respiratory uncouplers dinitrophenol [208, 209, 212] and carbonyl cyanide m-chlorophenylhydrazone [210–212], and abolished also by respiratory inhibitors sodium azide [208–210] and cyanide [212]. This inhibition of PTS activity was absent during anaerobic conditions [208, 209, 212], and during incubation with a substrate that the cell was not grown on (and thus was not configured to metabolise), leading to the suggestion that it is in fact the energisation of the plasma membrane during respiration that results in the inhibition of the PTS, and as such, a physical regulatory mechanism has been proposed which is dependent on proton motive force [213]. The inhibition of PTS activity under respiratory conditions was more pronounced in *atp* mutants, suggesting once more that the inability of the cell to dissipate the pmf by ATP generation resulted in more severe inhibition of the PTS [211]. Similarly, aerobically growing *E. coli* mutants expressing increasing concentrations of the soluble F_1 ATPase domain were shown to increase glucose oxidation (and thus glucose uptake) in the study by Koebmann *et al.* [101]. It has also been shown that in fact α-MG uptake is at least partially inhibited by reducing the intracellular ATP concentration with arsenate [214] and thus PTS activity is not merely dependent on respiratory activity uncoupled from ATP concentration. Indeed, altering the free energy state of the cell ($\frac{[ATP]}{[ADP]}$) using either *atp* mutants with varying ATP synthase expression levels, metabolic uncouplers (DNP), or even respiratory inhibitors (KCN) has shown unequivocally that PTS uptake activity is directly correlated with the $\frac{[ATP]}{[ADP]}$ ratio, and not necessarily with the pmf [118] when grown aerobically on succinate. That is, a decrease in ATP synthase activity, reduces the $\frac{[ATP]}{[ADP]}$ ratio and is correlated with reduced α-MG uptake by the PTS. The $\frac{[ATP]}{[ADP]}$ ratio (which is log-linearly related to phosphorylation potential, ΔG_p), is thus argued to be the thermodynamic driving force behind glucose uptake. In contrast, the *atp* mutant study by Koebmann *et al.* [101], demonstrated increased glycolytic flux upon lowered $\frac{[ATP]}{[ADP]}$ ratio, and thus control of glycolytic flux by the demand for ATP. Additionally, growing *E. coli* cells aerobically in the presence of weak acids—which uncouple respiration by lowering intracellular pH forcing the cell to expend ATP in

pumping protons out of the cytoplasm, lowering the $\frac{[ATP]}{[ADP]}$ ratio–generally resulted in a decrease in growth rate, with no change in glucose uptake [215]. This suggests that in fact that the $\frac{[ATP]}{[ADP]}$ ratio has little effect on glucose uptake, and as it is the demand reactions (i.e. growth rate) which are displaced in response to ATP dissipation, that the supply block exercises control over glycolytic flux to ATP. PEP and PYR levels show no correlation with PTS uptake rates [216], however total inhibition of PEP formation in membrane vesicles naturally inhibited PTS activity [212].

Though complex and somewhat contradictory, there are a number of clear trends in these studies. Under aerobic conditions, in the majority, respiratory activity is negatively correlated with PTS activity. Though it is not clear whether the PTS activity depends solely upon respiratory activity, or the energy charge of the cell, or to some degree on both of these. What is clear however, is that under anaerobic conditions where a low pmf is generated, inhibiting or uncoupling respiration has little to no effect on PTS activity.

E. coli has a diverse set of channels for carbon flux allowing it to balance both respiration and fermentation to match varying O_2 availability, as well as balancing fermentation products to achieve redox balance for growth on substrates with different oxidation states (reviewed in [217]). The mixed acid fermentation pathway in *E. coli* produces varying ratios of acetate, ethanol, lactate, succinate and α-ketoglutarate (Fig. 4.1). Glycolysis of one molecule of glucose produces two reducing equivalents, and as such, homolactic fermentation or a balance of equal levels of acetate and ethanol fermentation will maintain redox balance stoichiometrically (the acetate pathway does not consume any NADH and the ethanol pathway consumes 2 NADH, the equivalent of 2 reducing equivalents oxidised per glucose molecule). Redox balancing thus constrains the range of possible fermentation products and predominates as a determining factor of fermentative flux distribution. Aerobically grown phosphotransacetylase mutants Δpta do not redirect "overflow" flux through fermentation pathways, but rather, being constrained by the need to maintain redox balance, accumulate and excrete pyruvate [218].

Typical glycolytic pathways produce a net value of 2 ATP molecules per molecule of glucose by substrate-level fermentation, with the possibility of an additional ATP generated in *E. coli* per acetate molecule produced (Fig. 4.1). Cells importing glucose using the PTS (reviewed in [69]) use PEP as the phosphoryl donor to generate G6P rather than ATP, which is used by many other organisms (the PTS is a bacterial pathway and has not been shown to be present in the Archaea, the eukaryotes and is absent from a significant number of bacterial species [219]). Aerobic respiration produces significantly more ATP ($\pm$ 30) per glucose molecule and is also modular in that it consists of various dehydrogenases which feed electrons into a quinone pool in the plasma membrane from numerous respiratory substrates, pumping protons across the membrane, and thereby generating a proton motive force. The variance in ATP production by fermentation or respiration is thus dependent on the precise fermentative or respiratory configuration present, which in turn is determined primarily by the redox demand, O_2 availability and growth substrate.

Several glycolytic regulatory elements are present in *E. coli*. FBP significantly activates pyru-

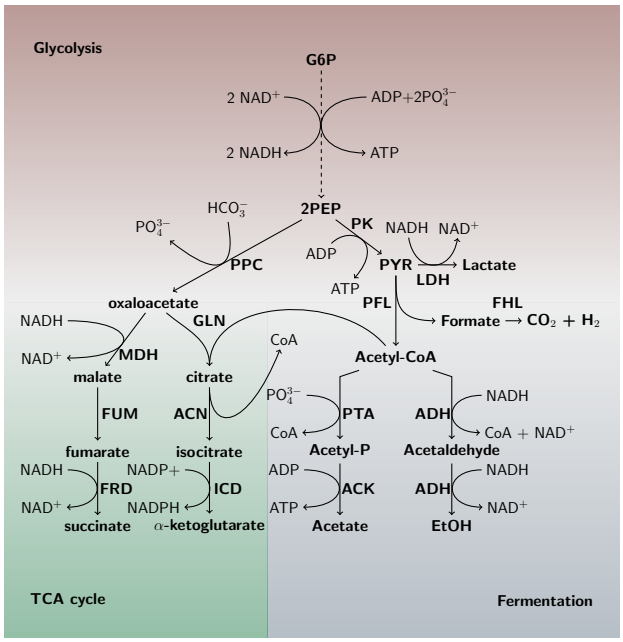


Figure 4.1: Mixed acid fermentation in *E. coli*, illustrating the alternate fates of glycolytic PEP and the topology of the important PEP-oxaloacetate-pyruvate node (reviewed in [122]). At high glucose concentrations and under low O_2 availability succinate dehydrogenase, succinyl-CoA synthetase and α-ketoglutarate dehydrogenase are repressed producing a branched version of the TCA cycle (green region) which produces biosynthetic precursors, succinate and α-ketoglutarate. Mixed fermentation (blue region) replenishes the NAD^+ pool for glycolysis (red region) to proceed. LDH - lactate dehydrogenase, PFL - pyruvate-formate lyase, PTA - phosphotransacetylase, ACK - acetate kinase, ADH - alcohol dehydrogenase, MDH - malate dehydrogenase, FUM - fumarase, FRD - fumarate reductase, GLN - citrate synthase, ACN - citrate hydro-lyase, ICD - isocitrate dehydrogenase, FHL - formate-hydrogen lyase, PO_4^{3-} - orthophosphate.

vate kinase I in a feed-forward manner, and transforms the rate characteristic from a sigmoidal shape to a simple Michaelis-Menten saturation shape [90, 220, 221], thus abolishing the positive cooperativity of PEP binding. Conversely, PEP exhibits a vast range of feedback inhibition on the first half of glycolysis including PGI, PFK and ALD [187, 222], and is known to inhibit TPI in other organisms (e.g. in yeast [223]). PEP is also the phosphate donor for the typical glucose transport system (the PTS). Additionally both PFK and PK are allosterically affected by ATP and ADP in a complex set of metabolite interrelationships [187, 224, 225]. Structurally therefore, there are at least four levels of allosteric regulation built into the glycolytic pathway of *E. coli*:

1. FBP accumulation in the early stages of glycolysis activates the later stage, ostensibly signalling the presence of substrate and possibly acting as a flux sensor for a transcriptional response [153].

2. PEP accumulation in the latter half of glycolysis inhibits the earlier preparative stage, possibly signalling saturation of the respiratory or fermentation pathways.

3. PEP is also required for glucose transport using the PTS system, a lack of late stage PEP may indicate low flux and thus an inability to process additional glucose.

4. And finally, the energy charge or $\frac{[\mathrm{ATP}]}{[\mathrm{ADP}]}$ ratio exhibits both positive and negative control on PFK and PK under various conditions dependent upon the concentrations of certain key metabolites, as well as being positively correlated with PTS activity [118].

Additionally it has been argued that FBP acts as a "flux sensor" of glycolysis in *E. coli* by its feed forward interaction with PK and inhibitory effect in conjunction with the Cra transcription factor which represses transcription of both *pfka* and *pykf* (and most other glycolytic genes) [226], thereby providing a "bridge" between metabolic and genetic control of glycolysis that specifically responds to flux [153].

To understand the behaviour and regulation of glycolysis, a kinetic model of glycolysis in *E. coli* was assembled out of three independent sub-models. The first model, representing the PTS which transports glucose into the cell, was developed by [68], and updated to reflect current experimental conditions (Section 1.2.2). The second model, described in Chapter 3, includes the first two reactions of glycolysis, PGI and PFK, and was employed as described. The third model representing the bottom half of glycolysis, from ALD to lactate dehydrogenase (LDH), was modelled in a similar fashion using *in vitro* NMR time courses of enzymatic reactions, with the exception that instead of using spline-fitting to estimate reaction rate of each reaction individually as in Chapter 3, the lower glycolytic model was fitted as a whole to a set of aggregated reaction series. Several variations on the model of the bottom half of glycolysis were constructed and fitted to *in vitro* datasets (e.g. including combinations of cytochrome-b$_d$ [cytb$_d$], adenylate kinase [ADK] and phosphoenolpyruvate synthetase [PPS]), incorporating various literature parameter values where possible. In this way comparisons could be made between primarily *literature parameterised models* and primarily *fitted parameter models*, and between models with various background activities

affecting, amongst others, the $\frac{[ATP]}{[ADP]}$ and $\frac{[NADH]}{[NAD^+]}$ ratios. Validation of the modelling results was achieved by comparing model behaviour to:

1. *in vivo* fermentation flux outputs of the bacterium determined using ^{13}C NMR,
2. *in situ* behaviour of permeabilised cells under similar conditions to the *in vitro* enzyme assays monitored using ^{31}P NMR,
3. literature steady state internal metabolite concentrations, and
4. *in vitro* internal metabolite accumulation of key intermediates as observed using ^{31}P NMR.

The following sections describe the construction and validation of these models. For a natural presentation experimental details are presented in the respective sections and are not collected into a separate section.

4.2 Model construction: lower glycolysis

To generate enzymatic reaction data for model-fitting of the lower half of glycolysis, a variation on the technique introduced in Chapter 3 was employed which was deemed necessary on account of the complex interdependency of the pathway constituents and the invisibility of unphosphorylated metabolites below PEP to ^{31}P NMR. *In vitro* ^{31}P NMR time series were collected as before, using cell extracts and various starting substrate/product/modifier concentrations. However, instead of approximating the time series data with splines and fitting an individual rate equation to paired metabolite/rate data (as in Chapter 3), a model of lower glycolysis representing all of the reactions present in the *in vitro* extracts was constructed and fitted to the reaction data by simulating the reactions under identical starting conditions. The key motivation for modelling lower glycolysis as a whole rather than individually as separate reactions is the interdependence of the reactions; for example, it is cumbersome if not impossible to distinguish between the activity of phosphoglycerate kinase (PGK) and pyruvate kinase (PK) when starting a reaction time series with metabolites in the lower half of glycolysis (e.g. 3PG), as both enzyme alter the $\frac{[ATP]}{[ADP]}$ ratio and metabolites appearing after PK (e.g. pyruvate and the fermentation products) are not visible to ^{31}P NMR.

For growth conditions and extraction methods refer to Section 3.3.

4.2.1 NMR spectrometry of reaction time series

All reaction components were from Sigma-Aldrich and prepared in 100 mM PIPES (pH 7.2, corrected by the addition of 10 M NaOH). Triethyl phosphate (TEP, usually at 5 or 10 mM final concentration) was introduced as an internal standard due to its metabolic inertness [205]. A typical assay in a 5 mm glass tube was composed of: 50 µmol TEP, 100 µl D_2O, 100 µl cell extract, 100 µl per substrate/product/modifier/co-factor, 5 µmol Mg^{2+}, and filled to a final volume of 1 ml

with 100 mM PIPES (pH 7.2). A blank (excluding cell extract) was used to tune the spectrometer, acquire a lock signal and shim the instrument before the reaction was started by removing the tube from the instrument, adding 100 µl cell extract, mixing several times by inversion, and re-inserting the tube. Data acquisition was initiated once a stable lock signal was acquired.

^{31}P NMR was performed at 25 °C with the probe spinning at 10 Hz, a pulse angle of 90°, a repetition time of either 1.5 s (1.0 s acquisition, 0.5 s relaxation) or 1.3 s (0.8 s acquisition, 0.5 s relaxation) to collect 64 - 128 transients, on 3 different Varian NMR spectrometers:

1. A 600 MHz spectrometer operating at 242.87 MHz
2. A 400 MHz spectrometer operating at 161.89 MHz
3. A 300 MHz spectrometer operating at 121.33 MHz

Acquisition was performed with proton decoupling (Waltz-16) on, including the Nuclear Overhauser Effect (NOE) for signal enhancement (to achieve this, decoupling is additionally performed during the relaxation delay [151]). In rapid-sampling NMR it is typically not possible to accommodate the full spin-lattice relaxation (T1) of the nucleus of interest. ^{31}P T1 relaxation times were determined for typical metabolites present in this study and ranged from 0.2 s (nucleoside phosphates) to 6 s (TEP). Species concentrations were thus calibrated for incomplete relaxation and differences in NOE enhancement with a fully-relaxed spectrum with NOE suppressed (30 s relaxation time, 5 × the longest T1) of either an equilibrated enzymatic reaction assay mixture, or a cocktail of metabolites of interest for reactions with high K_{eq}s.

The co-factor Zn^{2+} and reducing agent DTT which are often used to ensure ALD activity were excluded from the reaction mixture, thus bounding *in vitro* reactions to the enzymes below ALD (that is, from TPI onwards Fig. 4.3a).

4.2.2 Data processing

Collected reaction series data were processed using the NMRPy module as described in Chapter 2. Data were zero-filled 3 times and apodised by applying a 1 Hz exponential function to the FIDs. After Fourier-transformation, spectra were phased using the NMRPy.phase_auto() function which uses a Levenberg-Marquardt least-squares algorithm [168] to minimise total area under the peaks. To quantify metabolites, spectra were deconvoluted using a Lorentzian lineshape (NMRPy.deconv()) and spline-fitted when necessary to reduce noise. Note that, fitted splines were not used to approximate reaction rate as in Chapter 3, but merely to reduce noise. The parameter-fitting procedure employed is described in the following section. Data are plotted in Fig. 4.2.

4.2.3 Metaheuristic parameter-searching algorithms

Fitting models with numerous parameters to a relatively complex multidimensional dataset, as described in this work, presents an optimisation problem that cannot be solved using a local fitting

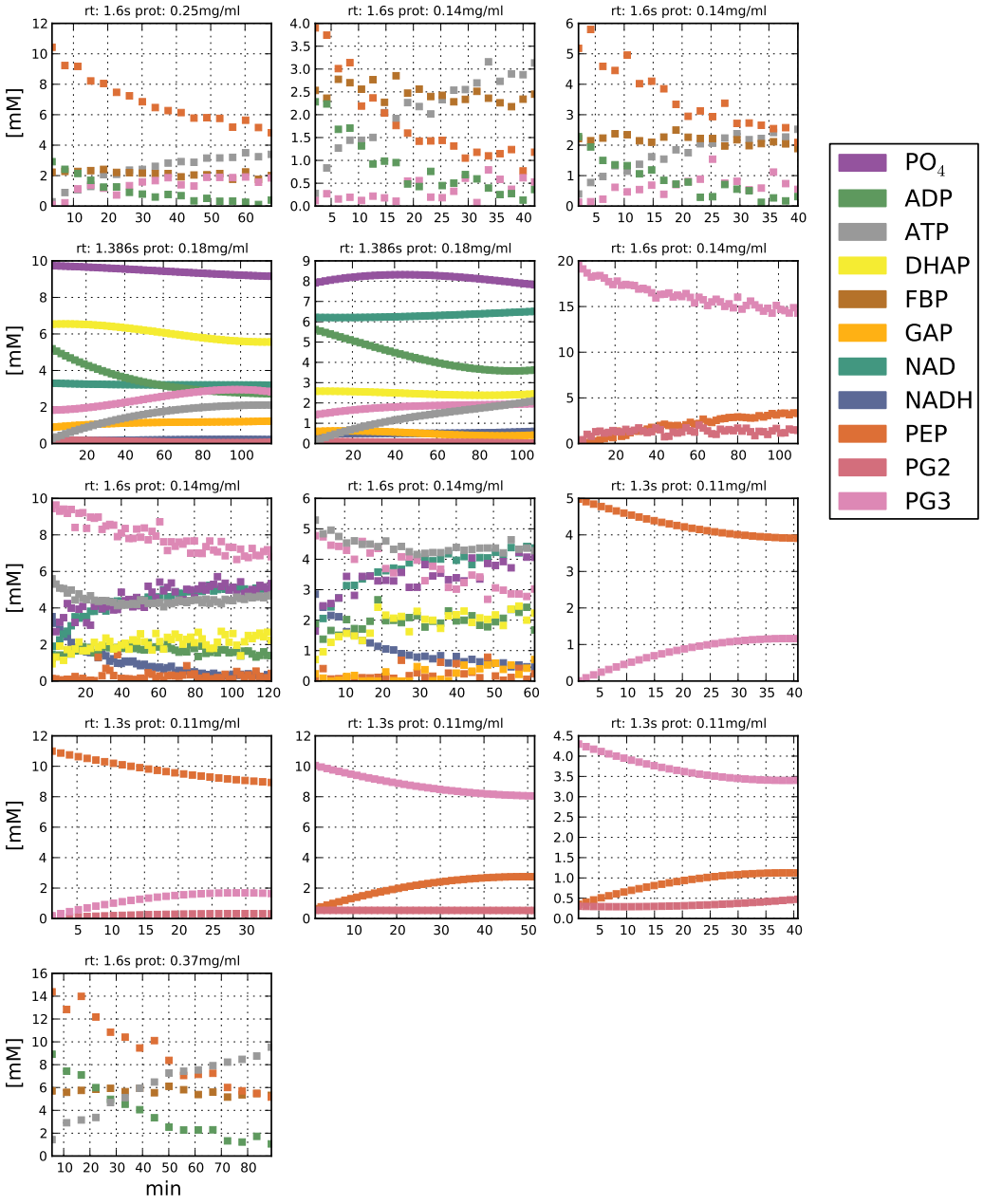


Figure 4.2: Time series of enzymatic reactions of the lower half of glycolysis acquired using ^{31}P NMR. Reaction time series were acquired with various starting concentrations of substrates, products, allosteric modifiers and co-factors as described in the text. FBP was not fitted. rt - repetition time, prot - protein concentration as determined by Bradford assay (Section 3.3).

algorithm like the popular least-squares methods. Many methods have been developed to search a particular "parameter space" for a global optimum, often taking inspiration from nature (reviewed in [227]). The task of a metaheuristic algorithm is to balance the ability to search wide areas of a "fitness landscape" without getting stuck in local minima, against the need to identify a particular local candidate solution and not remain "mobile" in parameter space, missing possible solutions.

Three algorithms were tested during the construction of the current model. Two evolutionary algorithms: a Genetic Algorithm [228] and a Differential Evolution Algorithm [229], were used. These are population based metaheuristic algorithms that are modelled on the process of Darwinian evolution. The third algorithm tested was a Particle Swarm Optimisation Algorithm [230, 231], modelled on the cooperative behaviour of swarming organisms.

- **Genetic algorithms** (GAs) are directly allegorical of evolutionary biology. A population of candidate solutions called a "chromosome" (each comprised of a particular selection of parameter values called "genes") is subjected to iterative selection based on a fitness criterion (e.g. Residual Sum of Squares), and a particular percentage of the fittest individual chromosomes are allowed to recombine to produce a "filial generation" [228]. Members of the following generation are mutated randomly at a specified rate. The fitness of the population increases progressively until an optimum is reached or some other specified termination criterion is achieved (e.g. number of generations, Residual Sum of Squares). Mutation and cross-over rates, and population size are specified beforehand and can play a large role in the efficacy and efficiency of the algorithm. Several modifications have been proposed to improve the efficiency of the GA such as variable mutation and cross-over rates [232], elitist selection, parallel computation with migration between nodes, linkage-learning which identifies and preserves beneficial epistatic relationships between specific genes [233] and many others.

- **Differential evolution algorithms** (DEs) work by recombination of randomly selected candidate solutions, or "agents" in solution space. Typically an agent x is randomly selected, and a putative new position in solution space is calculated by recombining three other randomly selected agents in the population according to a simple mathematical formula. If the new position provides more fitness than the current position of x, the position of x is updated [229]. This process is iterated until termination criteria are met.

- **Particle swarming optimisation algorithms** (PSOs) typically involve a swarm of particles in solution space that are moved around with trajectories informed by their own historically best positions (in terms of "fitness"), and the swarm's global best historical position. PSOs have a very high efficiency for relatively few candidate solutions by virtue of the simultaneous optimisation of each candidate with respect to its own history and that of the swarm. Ultimately the swarm will converge upon a particular location in parameter space identifying it as the solution.

Invariably, the PSO and DE algorithms converged on a solution sooner than the GA, however,

the GA was able to find better solutions. As such, the GA was selected for initial parameter estimation and `Python`'s `multiprocessing` package (docs.python.org) was employed to reduce runtime through process parallelisation. Parameter fitting was performed on a cluster provided by the Centre for High Performance Computing (www.chpc.ac.za). Parameter estimation using the GA is described below.

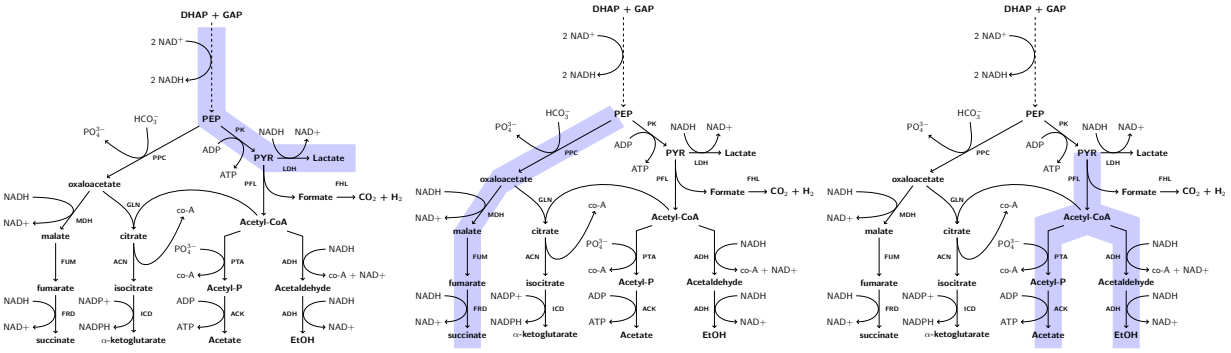
4.2.4 Parameter estimation by model fitting to *in vitro* NMR time course data

During the construction of the lower glycolytic model, aggregated *in vitro* NMR reaction time courses were fitted with a series of sub-models representing the enzymes of the bottom half of glycolysis (from TPI to LDH). To collate the various reaction time series during model fitting, repetition times were normalised by protein concentration. ^{31}P NMR is unable to detect glycolytic intermediates once they have been dephosphorylated, and thus the fermentation reaction intermediates below PEP in the pathway (Fig. 1.1) were invisible to the current assay system, and the precise configuration of fermentation pathways unknown. As such, a variegated strategy was adopted: a series of models was constructed by "attaching" different combinations of the succinate fermentation pathway, a cytb$_d$ respiratory NADH-consumption reaction, the PPS reaction, and ADK to a model of the lower glycolytic enzymes and lactate fermentation. Fig. 4.3 displays the two fermentation pathways included in the model: succinate fermentation – a pathway produced by the branching of the TCA cycle under anaerobic and high glucose conditions, and lactate fermentation by LDH. The acetate/ethanol fermentation branch which is fed by PFL was excluded due to the high oxygen sensitivity of PFL [129], which is denatured unless maintained under strictly anaerobic conditions (a precaution intentionally not adopted in this study to reduce the complexity of reactions involving unphosphorylated intermediates below PK which are invisible to ^{31}P NMR). PPS was included in some models as a possible indicator of gluconeogenic activity due to the cells being harvested in late log phase and the accumulation of 3-carbon fermentation products, both factors which up-regulate PPS expression [129, 234]. The background NADH consumption reaction was included to account for the presence of the lysed membrane fraction in cell extracts, and thus a possible respiratory reaction between the NADH-consuming oxidoreductases NDH-I and NDH-II [235] and the high-affinity cytb$_d$ which is specifically present under microaerobic conditions [129, 132].

Finally, two types of fitting regime were adopted. In the first approach, literature kinetic parameters for the enzymes were included wherever possible and only those parameters for which there were no published literature values, were fitted. In the second approach almost all the parameters of the central glycolytic reactions were fitted (with the exception of maximal rates). Literature kinetic parameters were retained for enzymes in the fermentation pathways as no data for these reactions could be captured using ^{31}P NMR. Literature parameters and rate equations are summarised in Section 4.5.

As LDH represents an important fermentation outlet for *E. coli*, the maximal rate and pyruvate

binding constant were determined using a microtiter plate assay with the following constituents (Fig. 4.4): sodium pyruvate – 1 - 20 mM, NADH – 0.2 mM, 100 mM PIPES buffer pH 7.2 to a final volume of 100 µl (assays performed in triplicate, refer to Section 3.3.5 for additional details). Resulting kinetics agree well with literature: V_{max} = 1.3 µmol/min/mg (1.1 in both [150] and [236]), $PYR_{0.5}$ = 2.51 mM (2.5–5.0 depending on pH [237]). Additionally, significant cooperativity towards pyruvate has been demonstrated [237].



(a) (b) (c)

Figure 4.3: Three fermentation branches (shaded) of lower glycolysis. Glycolytic enzymes are excluded for brevity and are represented by the dashed line. Branches (a) and (b) which represent lactate fermentation and the succinate fermentation pathway (excluding PCK) were combined in various ways with PPS, ADK and an NADH-consuming reaction to fit to *in vitro* data. (c) The acetate/ethanol fermentation pathway was excluded from model-fitting to *in vitro* data due to the highly oxygen-labile PFL enzyme.

4.2.5 Model fitting: reassortment to generate putative models of *in vitro* experiments

Using various combinations of sub-model pathways and employing either a maximum or minimum number of literature parameters, 18 models were generated for fitting to the *in vitro* time series data and are tabulated in Table 4.1. "SUCC" refers to the enzymes: PEP carboxylase (PPC), PEP carboxykinase (PCK), malate dehydrogenase (MDH), fumarase (FUM) and fumarate reductase (FRD) of the succinate fermentation pathway. The first reaction, PPC, requires bicarbonate (HCO_3^-, Fig. 4.3b) which is typically produced by the carbonic anhydrase reaction either by conversion of dissolved atmospheric CO_2, or conversion of CO_2 produced by the reaction catalysed by formate hydrogenlyase (FHL) under non-respiratory conditions (Fig. 4.3c). For model fitting, HCO_3^- was assumed to be saturating and simply excluded from rate equations. It is important to note that fitted models including the succinate pathway and/or the ethanol/acetate pathway are *a priori* unlikely to be a true reflection of *in vitro* conditions due to the highly oxygen-sensitive PFL and FHL enzymes which are invariably denatured and repressed respectively upon cell lysis

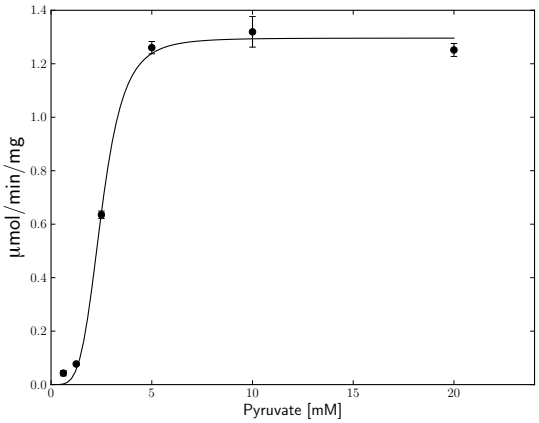


Figure 4.4: A coupled enzyme assay of LDH. LDH was assayed at various concentrations of pyruvate and initial rates were fitted with a uni-substrate Hill equation ($v = V_{max} \frac{PYR^h}{PYR^h_{0.5} + PYR^h}$) to estimate parameters. Fitted parameters are: $V_{max} = 1.3 \pm 0.03$ µmol/min/mg protein, $PYR_{0.5} = 2.51 \pm 0.08$, $h = 4.45 \pm 0.87$. Error is SEM. Cell extract was serially diluted until it was clear that initial rates were not being underestimated due to rapid reaction, and V_{max} was adjusted accordingly.

[129, 238, 239], which results in the concomitant lack of HCO_3^- necessary for the succinate pathway (PPC) (*E. coli* growing under low O_2 conditions are amply supplied with a high partial pressure of CO_2 [240]). At the least, HCO_3^- concentrations would be low enough not to produce detectable changes under the current experimental conditions due to the low hydration equilibrium constant of carbonic acid ($\frac{H_2CO_3(eq)}{CO_2(eq)} = 1.7 \times 10^{-3}$ at 298 K [241]) and the low mole fraction of dissolved CO_2 under atmospheric conditions[1]. The half-saturation constant of PPC is high when not activated by acetyl-CoA (ranging from 15 mM in the absence of acetyl-CoA to 0.35 mM in the presence of acetyl-CoA). With the disruption of PFL, no significant acetyl-CoA would be generated. Together, these considerations strongly suggest *a priori* that neither the succinate nor the ethanol/acetate fermentation pathways are active *in vivo* and the data should be adequately described by a simple homolactic fermentation model. However, for thoroughness and due to the fact that fermentation pathway metabolites below PEP are not visible to ^{31}P NMR, and thus the degree of inactivation of PFL was not known, a range of models including the unlikely fermentation pathways and excluding these pathways was fitted. Also, the bulkier models including the succinate pathway are more likely to be rejected by the Akaike criterion used below and they thus provide a comparative reference for the simpler, parsimonious models. Indeed the decision to use 18 model permutations was purposefully excessive to provide a more nuanced and thorough understanding of the distribution of Akaike scores.

[1]Henry's Law [242] predicts a dissolved CO_2 of about 0.5 mM under atmospheric partial pressure of CO_2 at 298 K (Henry's constant of 29.4 $\frac{\text{L.atm}}{\text{mol}}$), and thus a maximum of $1.7 \times 10^{-3} \times 0.5$ mM = 0.85 µM. This is significantly lower than the HCO_3^- binding constant for PPC (0.1 → 0.15 mM [243, 244])

A genetic algorithm (GA) was constructed in `Python` in which each parameter being fitted was represented as a "gene" on a "chromosome" in a population of "chromosomes". The fitting routine requires boundary values for each parameter, and a starting "generation" of chromosomes is instantiated by randomly populating with chromosomes; each gene (or parameter) was randomly sampled from a uniform distribution between specified parameter boundaries. Typically population size was 30. In each generation, all the chromosomes in the population were assigned a fitness score which involved using the parameters in a particular chromosome to parameterise the model being fitted, and a time simulation of the model was performed using the starting conditions of the 13 ^{31}P NMR datasets; fitness was calculated as the Residual Sum of Squares (RSS) between the simulation and the data. The "fittest" 10% of chromosomes in a particular generation[2] were allowed to recombine and mutate (initially at static specified rates, typically 0.8 and 0.2 respectively) to populate a new generation. This process was iterated 600 times. It was found that reducing the mutation rate sigmoidally from 90 % to 10 % over the course of the fitting iterations allowed the GA to identify a global solution with greater reliability; a strategy that has been suggested before [245, 246].

Putative parameter values produced using the GA were used as starting values for the Levenberg-Marquardt (LM) least-squares algorithm [168] to fine-tune parameter estimates, and to identify parameters that were not constrained by the data. Additionally, the LM algorithm outputs an estimated covariance matrix for the fitted parameter set which can be used to generate error estimates (error is calculated as $\frac{RSS}{Deg.of Freedom}\sigma^2$, where σ^2 is the variance in the fitted function for a particular set of parameters). A transposition of a simulation of model O with fine-tuned fitted parameters is shown in Fig. 4.5.

4.2.6 Model selection: the Akaike information criterion

The process of selecting the best model for a particular application involves balancing explanatory power against the number of *a priori* assumptions required for the model [249]. This parsimony principle famously known as Occam's Razor, argues that one should make as few assumptions as possible in the process of explaining a particular phenomenon, or more accurately, amongst competing hypotheses priority should be given to the ones with the fewest assumptions [250]. This trade off between explanatory power and simplicity has been quantified in a metric known as the Akaike Information Criterion (AIC) [251]. In statistical terms, when fitting models to data, simplifying a model by reducing the number of parameters concomitantly increases bias (e.g. fitting a linear model to a parabolic dataset); enhancing the explanatory power of the model by increasing the number of parameters increases variance (e.g. fitting a parabola to a linear dataset). It is this trade-off between bias and variance (or precision) that is quantified by the AIC [249]. In information-theory terms, the AIC quantifies the relationship between the Kullback-Leibler measure of information loss and the maximum-likelihood of a particular model. Thus when comparing

[2] In this case "fitness" was minimised as it was represented by the RSS between model and data.

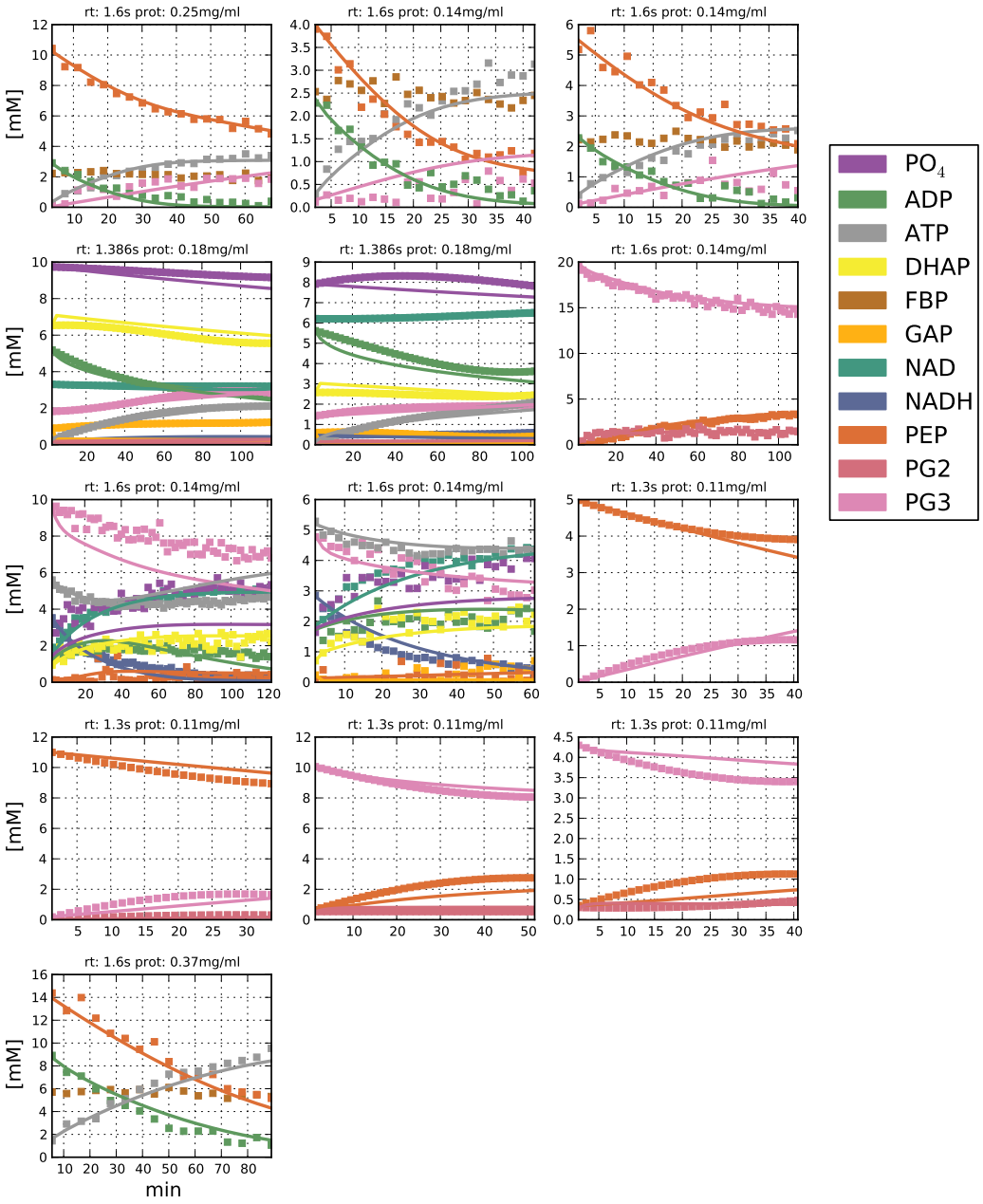


Figure 4.5: Model O fitted simulation transposed onto time series of enzymatic reactions of the lower half of glycolysis acquired using ^{31}P NMR. Reaction time series were acquired with various starting concentrations of substrates, products, allosteric modifiers and co-factors as described in the text. rt - repetition time, prot - protein concentration as determined by Bradford assay (Section 3.3). Data is identical to Fig. 4.2.

Table 4.1: Combinations of fermentative pathways used to construct putative sub-models for parameter-fitting of *in vitro* NMR reaction time course data. +/- indicate inclusion or exclusion of a particular model feature. See text for details.

	SUCC[a]	PPS[b]	ADK[c]	cyt_{bd}[d]	LIT[e]	
A	-	+	-	+	+	
B	-	+	-	+	-	
C	+	-	-	+	+	
D	+	-	-	+	-	
E	+	+	-	+	+	
F	+	+	-	+	-	
G	+	+	+	+	+	
H	+	+	+	+	-	
I	+	+	+	-	+	
J	+	+	+	-	-	
K	+	+	+	+	-	All Hill-coefficients fitted[f]
L	±	-	+	+	-	Δmal[g]
M	-	-	-	+	-	vf_{LDH} was allowed to vary.[h]
N	-	-	-	+	+	vf_{LDH} was allowed to vary.[h]
O	-	-	+	+	-	
P	-	-	-	+	-	
Q	-	-	+	+	+	
R	-	-	+	-	-	

[a] The succinate fermentation pathway include PPS, PCK, MAL, FUM and FRD. During fitting of the lower glycolytic kinetic parameters, the rate equations for these reactions (and that of PFL) were simplified to irreversible forms in order to reduce the number of parameters to be fitted. When the succinate pathway and PFL were finally included in the full model, reversible rate equations and unfitted literature parameters were employed (as indicated in Section 4.5).

[b] PEP synthetase

[c] Adenylate Kinase

[d] Background NADH consumption by the high-affinity microaerobic cytochrome b_d

[e] The distinction in this column is between those models with all the literature values included and those with very few literature values included.

[f] This includes the enzymes for which no cooperativity has been reported.

[g] Only PPC and PCK of succinate pathway, a futile cycle mediated by PPC and PCK has been identified [247, 248].

[h] LDH displays a number of cysteine groups near the surface of the protein which are subject to auto-oxidation which deactivates the enzyme. In this model Vf_{LDH} was fitted to compensate for possible deactivation in the coupled assay determination of maximal rate.

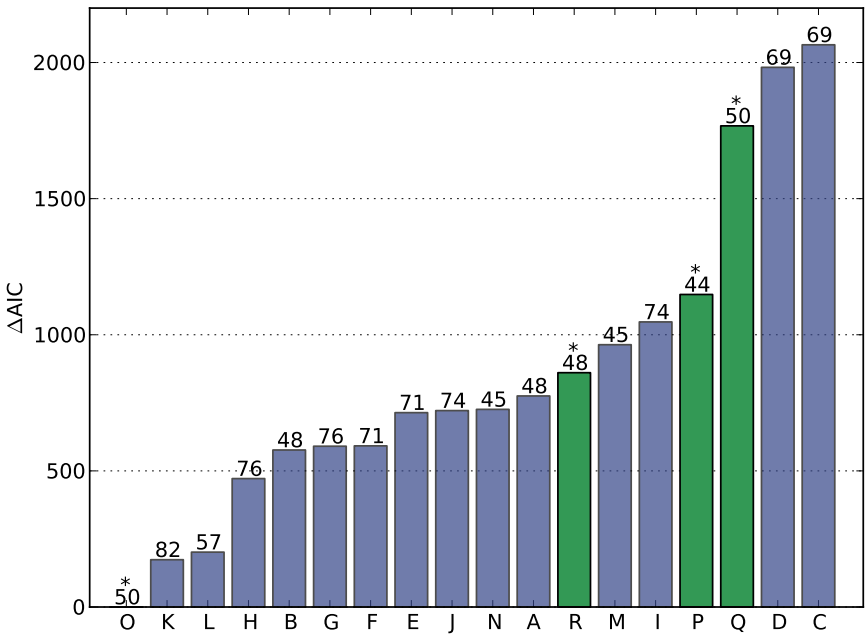


Figure 4.6: ΔAIC scores for 18 putative models after fitting to reaction time series. Calculated relative to model O. Parameter number $(K-1)$ is in superscript above the respective bar in the plot. Sample size (n), the total number of data points, was 2047. For model descriptions see Table 4.1. Simple homolactic models selected for further modelling are indicated by an asterisk.

various models, the AIC rewards minimisation of information loss and penalises for the number of parameters. For datasets in which the sample size is below 40 times the number of parameters ($\frac{n}{K} < 40$), the second-order corrected AIC is recommended [249, 252]:

- First-order: $AIC = n\log(RSS) + 2K$
- Second-order: $AIC = n\log(RSS) + 2K + \frac{2K(K+1)}{n-K-1}$

Where n = sample size, RSS = Residual Sum of Squares, and K = the number of parameters (note that when using a least-squares approach rather than a maximum-likelihood approach, variance is approximated and is included as a parameter, thus the value of K has to be incremented by one). When comparing various models using the AIC, a ΔAIC value is often reported, which is a relative measure calculated as the difference between the model under consideration and the best model in the set (i.e. with the lowest AIC score).

Fig. 4.6 shows the ΔAIC score calculated for each of the 18 putative models (Table 4.1) after fitting with the GA and the subsequent LM to the time series data in Fig. 4.2. It is clear from the ΔAIC scores that a minimum is achieved by the first 3 models. Models K and H represent "all-inclusive" fitting attempts in which the full succinate pathway, PPS, ADK and NADH consumption were included with as few literature parameters as possible. That is, almost all parameters were fitted (a minimal selection of literature values was included) and the models have high parameter numbers (82 and 76 respectively, in model K all the Hill coefficients were also fitted, explaining the discrepancy in parameter number). As expected by virtue of PFL disruption and the lack of bicarbonate for PPC, neither of these parameter-heavy models were able to perform as well as model O, a far simpler 50 parameter model including only ADK and NADH-consumption in addition to the default lactate fermentation pathway. This suggests, in conjunction with earlier biochemical considerations about HCO_3^- concentration and O_2 deactivation of PFL, that indeed simpler homolactic models offer a better description of the *in vitro* data. Note that the models with a high number of literature values, and thus a low number of fitted parameters, as expected, performed worse on the AIC scoring as they were subjected to higher degree of constraint which is not included in the quantification. That is, the AIC was calculated taking total parameter number into account, not simply fitted parameter number. Considering the results of Fig. 4.6—specifically that a simple homolactic model structure is sufficient to describe the behaviour of the data—a selection of homolactic models was made which excluded PPS activity and employed the experimentally determined maximal LDH activity. The resulting four models—O, P, Q and R—were used as the basis for further model development. The aggregated parameters for the four selected models after fitting are shown in Table 4.2 and in Fig. 4.7.

The primary locus of variation amongst the fits was around GAPDH and PGK. The GAPDH and PGK reactions exhibit relatively high equilibrium constants, favouring the reactions away from

BPG; that is, the GAPDH reaction is favoured in reverse ($K_eq = 0.0004$) and the PGK reaction in the forward direction ($K_{eq} = 4000$). Thus, a thermodynamic obstacle is only avoided by a constrained relationship between the kinetic parameters of these enzymes which is apparent during the fitting process as changes in the kinetic parameters of one of these enzymes results in significant changes in the parameters of the other. For example, Model P fitted a significantly lower GAP saturation constant for GAPDH, and consequently a significantly higher BPG saturation constant for PGK than the other models. All non-literature model fits (O, P and R) produced higher product binding constants than the literature values, suggesting that under the current experimental conditions less product-inhibition and thus greater activity of the enzymes was observed. It is noteworthy that GAPDH is the only redox-associated glycolytic enzyme. Thus, the particular thermodynamic and kinetic configuration around this enzyme is of global regulatory importance. The parameter fits for PGM and ENO also demonstrated a degree of variation.

4.2.7 *In situ* validation of the fitted model of lower glycolysis

To validate the ability of the fitted models O, P, Q and R to correctly predict dynamic intracellular metabolic behaviour of glycolytic intermediates under defined conditions, two *in situ* assays were performed using permeabilised cells and monitored using ^{31}P NMR, and the *in situ* progress curves were compared with model simulations under identical starting conditions. Flash frozen cells (3.0 mg/ml protein as determined by the Bradford protein assay [203]) were resuspended to 0.8 ml in 100 mM PIPES buffer (pH 7.2) and permeabilised with 0.5 % Triton X-100 (Roche), vortexing for 30 s and leaving at room temperature for 30 min before assaying, according to the protocol in [253]. Cell integrity was verified by microscope. Reaction mixtures were prepared and reactions initiated under two different starting conditions by the addition of permeabilised cells to assay mixture in a 5 mm NMR tube. ^{31}P NMR was performed on a 300 MHz Varian spectrometer (operating at a frequency of 121.33 MHz) at 25 °C, with a mild 5 Hz spinning of the tube to keep the cells mixed but prevent centrifugation. A repetition time of 1.5 s (1.0 s acquisition, 0.5 s relaxation) was used, and a 45° and 90° pulse angle for experiments 1 and 2 respectively, with proton decoupling (Waltz-16) and the NOE for signal enhancement (i.e. decoupling was on during acquisition and relaxation). 64 transients were collected per FID (1.6 min per FID). After equilibration of the reactions, a relaxation calibration curve with decoupling on and NOE excluded (i.e. decoupling on only during acquisition) was acquired with a 30 s relaxation delay. The following reaction mixes were prepared:

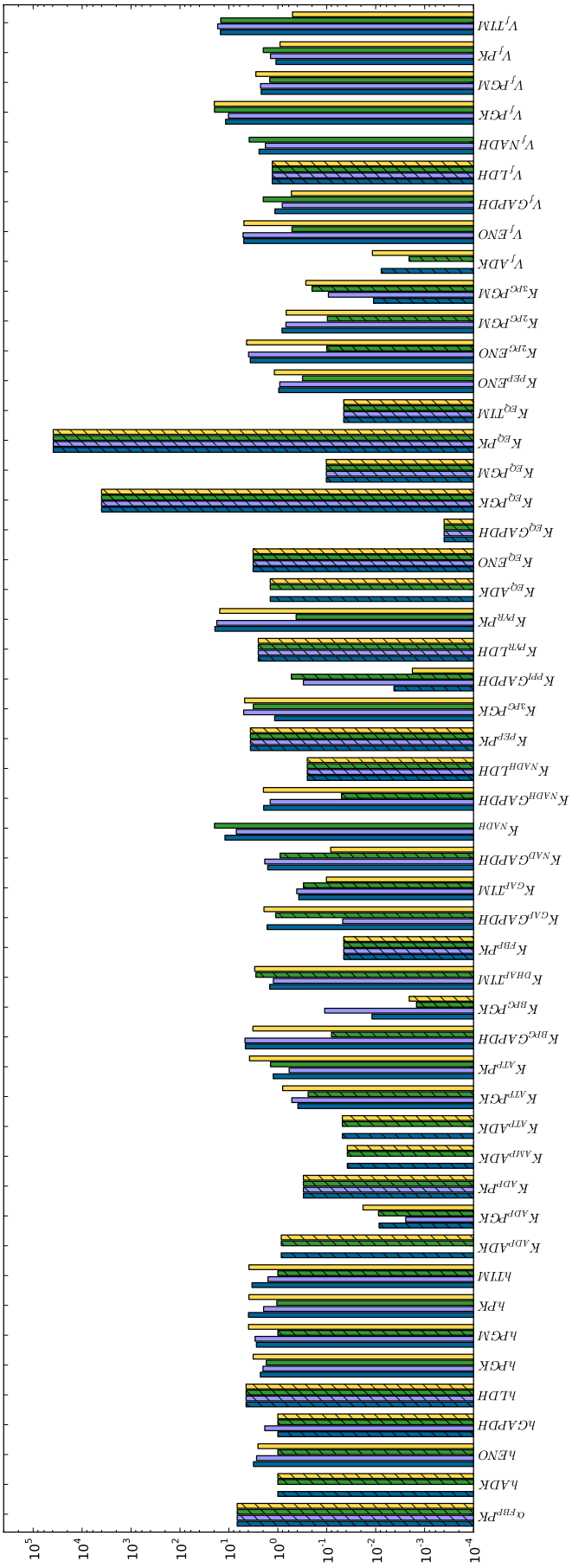


Figure 4.7: Parameters for sub-models O (blue), P (cyan), Q (green) and R (yellow) after fitting to NMR reaction time series data. Error bars as calculated from the LM fitting algorithm were too small to be visible and have thus been excluded. Hatched bars indicate unfitted literature parameters. Data are log-scaled for comparison. For a description of the sub-models refer to Table 4.1. Parameter values can be found in Table 4.2

Table 4.2: Parameters for sub-models O, P, Q and R after fitting. Unfitted parameters are indicated by the lack of an error estimate. Error is SEM as calculated by the Levenberg-Marquardt least-squares algorithm. References for unfitted parameters can be found in section 4.5.

	O	P	Q	R
$\alpha_{FBP}PK$	6.762	6.762	6.762	6.762
$hADK$	1.0		1.0	1.0
$hENO$	3.1412 ± 0.0093	2.717 ± 0.0535	1.0	2.5436 ± 0.0538
$hGAPDH$	1.0	1.8491 ± 0.039	1.0	1.0
$hLDH$	4.45	4.45	4.45	4.45
$hPGK$	2.2727 ± 0.0095	2.0124 ± 0.0717	1.7189 ± 0.0079	3.192 ± 0.059
$hPGM$	2.7353 ± 0.0096	2.9169 ± 0.1164	1.0	3.9686 ± 0.063
hPK	3.9767 ± 0.015	1.9509 ± 0.0526	1.0472 ± 0.0039	3.8961 ± 0.1099
$hTIM$	3.3662 ± 0.0112	1.5919 ± 0.0694	1.0	3.8858 ± 0.0876
$K_{ADP}ADK$	0.85		0.85	0.85
$K_{ADP}PGK$	0.0085387 ± 0.0	0.002432 ± 0.0001	0.008757 ± 0.0	0.018161 ± 0.0007
$K_{ADP}PK$	0.3	0.3	0.3	0.3
$K_{AMP}ADK$	0.038		0.038	0.038
$K_{ATP}ADK$	0.048		0.048	0.048
$K_{ATP}PGK$	0.38871 ± 0.0022	0.51765 ± 0.0167	0.24	0.79556 ± 0.0218
$K_{ATP}PK$	1.2423 ± 0.005	0.58682 ± 0.0202	1.4123 ± 0.0078	3.7898 ± 0.0732
$K_{BPG}GAPDH$	4.6028 ± 0.0152	4.688 ± 0.0554	0.08	3.2284 ± 0.0614
$K_{BPG}PGK$	0.011896 ± 0.0001	0.1102 ± 0.002	0.0014684 ± 0.0	0.0020559 ± 0.0
$K_{DHAP}TIM$	1.465 ± 0.0042	1.2337 ± 0.0403	2.8	2.9673 ± 0.0393
$K_{FBP}PK$	0.045	0.045	0.045	0.045
$K_{GAP}GAPDH$	1.6547 ± 0.0066	0.047258 ± 0.0019	1.1	1.9199 ± 0.0349
$K_{GAP}TIM$	0.37326 ± 0.0019	0.41035 ± 0.0121	0.3	0.10151 ± 0.0021
$K_{NAD}GAPDH$	1.61 ± 0.0054	1.8508 ± 0.0663	0.9	0.083012 ± 0.0028
K_{NADH}	12.034 ± 0.0524	7.0545 ± 0.1998	19.68 ± 0.0704	
$K_{NADH}GAPDH$	1.9447 ± 0.0104	1.4325 ± 0.0412	0.05	1.966 ± 0.0287
$K_{NADH}LDH$	0.25	0.25	0.25	0.25
$K_{PEP}PK$	3.63	3.63	3.63	3.63
$K_{3PG}PGK$	1.1603 ± 0.0036	4.9915 ± 0.1486	3.1781 ± 0.0122	4.7623 ± 0.0755
$K_{PPI}GAPDH$	0.0042439 ± 0.0	0.30036 ± 0.0085	0.53	0.0017805 ± 0.0001
$K_{PYR}LDH$	2.51	2.51	2.51	2.51
$K_{PYR}PK$	19.316 ± 0.1054	17.66 ± 0.6115	0.4224 ± 0.0019	15.393 ± 0.1417
$K_{eq}ADK$	1.4286		1.4286	1.4286
$K_{eq}ENO$	3.2	3.2	3.2	3.2
$K_{eq}GAPDH$	0.0004	0.0004	0.0004	0.0004
$K_{eq}PGK$	4000.0	4000.0	4000.0	4000.0
$K_{eq}PGM$	0.10204	0.10204	0.10204	0.10204
$K_{eq}PK$	$38,900$	$38,900$	$38,900$	$38,900$
$K_{eq}TIM$	0.045	0.045	0.045	0.045
$K_{PEP}ENO$	0.95904 ± 0.0026	0.91253 ± 0.0264	0.31296 ± 0.0012	1.1774 ± 0.0203
$K_{2PG}ENO$	3.6556 ± 0.0142	3.9623 ± 0.0588	0.1	4.3548 ± 0.0549
$K_{2PG}PGM$	0.82043 ± 0.0023	0.68339 ± 0.022	0.097	0.67479 ± 0.0114
$K_{3PG}PGM$	0.011071 ± 0.0	0.092608 ± 0.0029	0.2	0.26773 ± 0.007
V_fADK	0.0077096 ± 0.0		0.0020568 ± 0.0	0.011651 ± 0.0004
V_fENO	4.9658 ± 0.0185	5.1385 ± 0.122	0.51412 ± 0.0018	4.9359 ± 0.1141
V_fGAPDH	1.1458 ± 0.0042	0.81229 ± 0.0214	1.9858 ± 0.0051	0.52803 ± 0.0112
V_fLDH	1.3	1.3	1.3	1.3
V_fNADH	2.4205 ± 0.0092	1.7942 ± 0.0476	3.8454 ± 0.0121	
V_fPGK	11.67 ± 0.0391	10.21 ± 0.3324	19.684 ± 0.0794	19.891 ± 0.3958
V_fPGM	2.1858 ± 0.0076	2.2576 ± 0.0939	1.4605 ± 0.0083	2.8116 ± 0.0888
V_fPK	1.0906 ± 0.0035	1.411 ± 0.0354	1.9787 ± 0.007	0.9021 ± 0.0141
V_fTIM	14.873 ± 0.0442	16.963 ± 0.7453	14.691 ± 0.0224	0.50083 ± 0.0249

	Experiment 1	Experiment 2
TEP	5 mM	5 mM
D_2O	10%	10%
cells	350 µl (35%)	350 µl (35%)
G6P	7 mM	0
FBP	0	5 mM
ATP	4.0 mM	2.5 mM
ADP	2.0 mM	2.5 mM
PO_4^{3-}	12 mM	10 mM
MgCl	5 mM	5 mM
$ZnSO_4$	0.5 mM	0 mM
DTT	0 mM	0.5 mM
NAD	10 mM	2.5 mM
PIPES	85 mM	85 mM
Final vol.	1 ml	1 ml

The four putative models were augmented with the PGI and PFK kinetic equations generated in Chapter 3 (Table 3.1) and an ALD reaction with literature kinetic parameters (Section 4.5). To account for the potentially significant differences of the *in situ* experimental environment when compared to the *in vitro* environment used to generate the model [254–256], a number of additions and variations were made to the models to optimise the fit of the models to the data:

- the V_{max} of ALD was varied in Experiment 2 to account for the lack of supplemental Zn^{2+} in the assay buffer
- a non-specific ATP-consuming reaction was added to account for background ATP consumption (ATP $\rightarrow$ ADP + PO_4^{3-}, $v = k[ATP]$)
- DTT in experiment 2 was observed to reduce NAD^+ and as such a simple mass-action equation was included to account for this effect (DTT + $NAD^+ \rightarrow$ NADH, $v = k[NAD^+][DTT]$)
- background $cytb_d$ activity was varied to account for possible disruption of the NADH-consuming NDH-I/II—quinol—$cytb_d$ chain in the cell membrane by the detergent used in the permeabilisation process (this alteration had no effect on model R which lacks $cytb_d$ activity).

Though significant differences between *in vivo*, *in situ* and *in vitro* equilibrium constants, enzyme activities and allosteric properties have been observed [254–258], the only model parameters that were varied to account for these possible differences in the cellular environment were the total protein concentration and the equilibrium constant of the GAPDH reaction ($K_{eq}GAPDH$, which

responds to changes in ionic strength [259]). No fitted model parameters other than the $cytb_d$ activity were altered.

Results of the *in situ* time series of experiments 1 and 2, and the associated model simulations under identical starting conditions are plotted in Fig. 4.8. Experiment 1 begins with G6P at the start of glycolysis; experiment 2 begins further down the glycolytic chain with FBP. Bearing in mind that conditions *in situ* can be significantly different from the highly diluted *in vitro* conditions of the assay system [254, 257], the model simulations present a more than qualitative agreement with the acquired data. A comparison of the RSS (Fig. 4.8) of the model predictions and the *in situ* data show that model P (the low literature parameter model lacking ADK activity) and to a lesser extent model Q (the high literature model with ADK and NADH consumption) were best able to predict the *in situ* behaviour.

4.3 Adding glucose uptake and fermentation capabilities to the fitted model

This section describes the augmentation of the constructed glycolytic model with a model of the PTS for glucose uptake, and fermentation pathways.

4.3.1 *In vivo* glucose uptake flux

To determine steady-state flux, an *in vivo* ^{13}C glucose uptake assay was performed in triplicate using a 600 MHz Varian NMR spectrometer. Cells were cultivated for NMR under standard conditions (described in Section 3.3). 3 × 1 l microaerobic cultures were harvested in mid-log phase (OD 0.45) by centrifugation for 7 min at 5,000 rpm (∼4200 × g) and a temperature of 4 °C. Pellets were resuspended in 100 mM PIPES buffer (pH 7.2) to 50 ml as a washing step and centrifuged a second time at 4,500 rpm (∼5,600 × g) for 10 min at 4 °C using a desktop centrifuge. These pellets were resuspended in minimal volume (around 4 ml) of 100 mM PIPES (pH 7.2) of which 2.4 ml was included in a final assay volume of 3 ml with 0.3 ml of a 0.5 M mixture consisting of $\frac{1}{6}$D-glucose-1-^{13}C:$\frac{5}{6}$D-glucose-^{12}C (to a final concentration of 50 mM glucose) and 0.3 ml D_2O. ^{13}C NMR was performed with the samples in 10 mm glass tubes at 150.87 MHz using a 0.6 s repetition time (0.1 s acquisition, 0.5 s relaxation) and a 90° pulse angle with proton decoupling on to collect 100 transients per FID (1 min). Tubes were lightly spun (10 Hz) to discourage precipitation. Spectra were corrected for incomplete relaxation by acquiring a fully relaxed spectrum of a non-reacting standard (15 s relaxation time) and adjusting integrals accordingly. Samples were stored on ice after harvesting and allowed to equilibrate in the spectrometer until a stable lock signal was achieved, indicating temperature acclimation. Acquisition was started after glucose was added to the samples to initiate the reaction. Acquired spectra are shown in Fig. 4.9. Time series were fitted with splines to extract uptake rates (using the scipy.interpolate module [163]). Peaks were normalised to buffer concentration for quantification. Rates were normalised to protein concentration which was

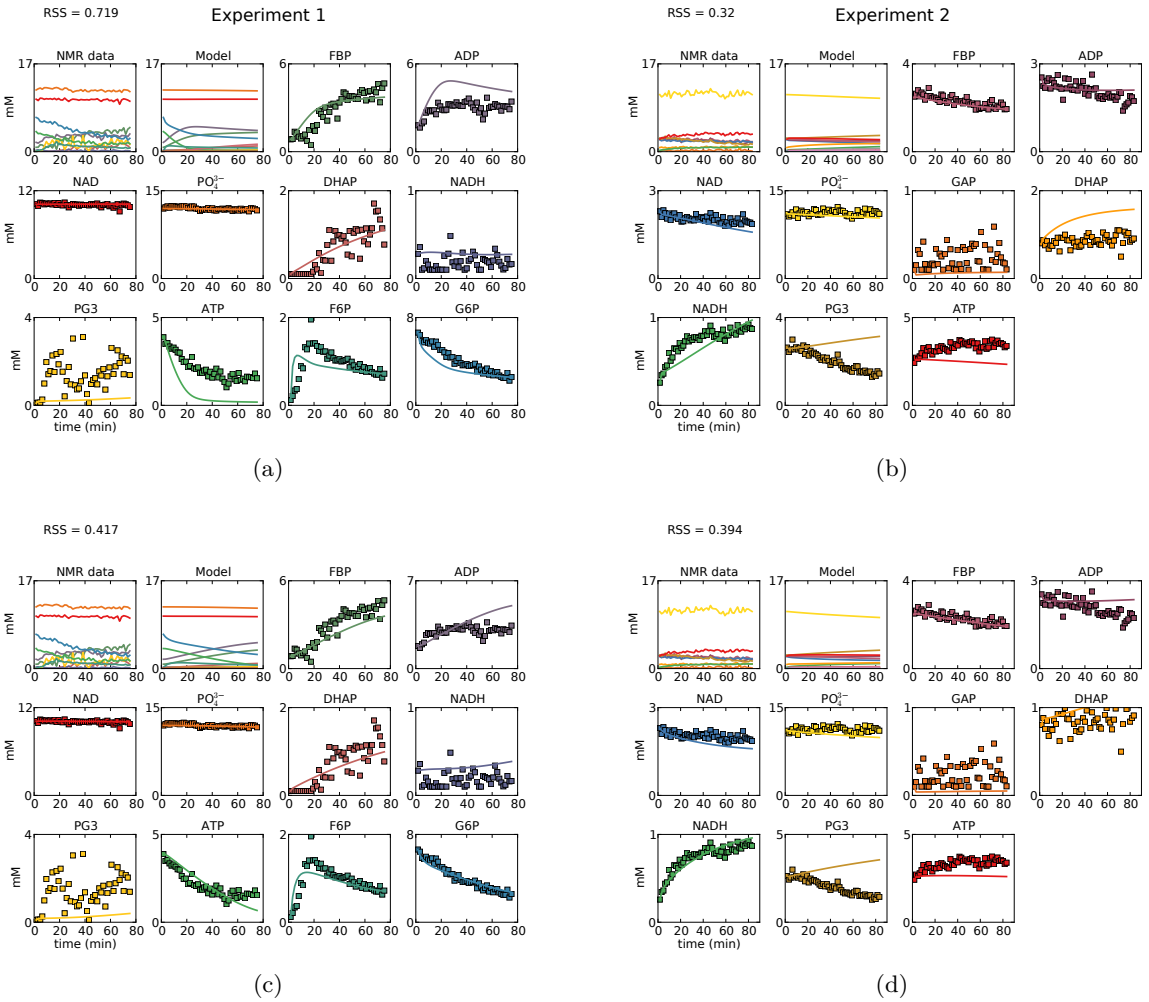
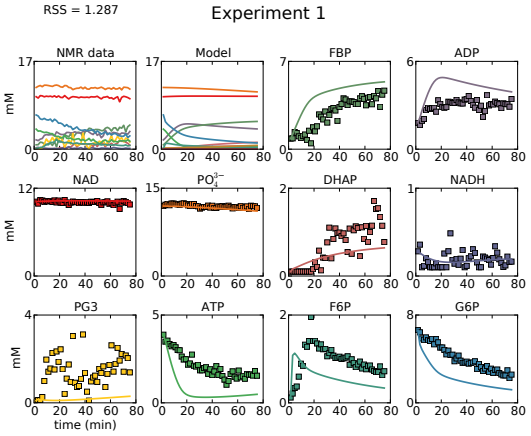
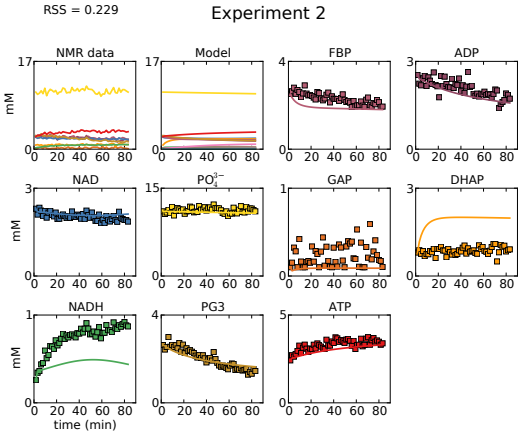


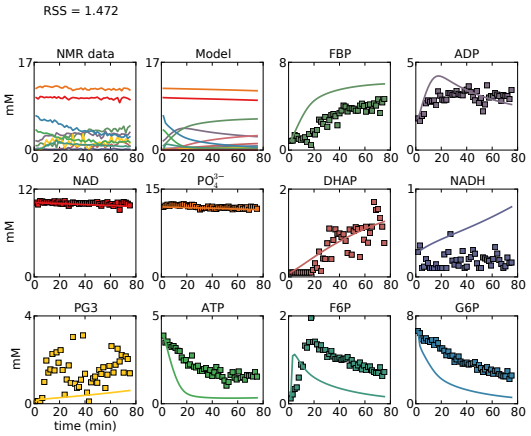
Figure 4.8: A comparison of *in situ* time series captured using ^{31}P NMR, and model simulations for selected homolactic models (a),(b) O – this model included ADK and background NADH consumption with few literature parameters, (c),(d) P – this model only included background NADH consumption with few literature parameters, (e),(f) Q – this model included ADK and background NADH consumption with mostly unfitted literature parameters, and (g),(h) R – this model included only the ADK reaction and excluded background NADH consumption, few literature values were used. The lines represent the model simulations; dots represent the *in situ* ^{31}P data. RSS values in the top left corner of each subfigure represent the Residual Sum of Squares between the model simulation and the *in situ* data and are scaled by the number of data points ($RSS = \sum_{i=1}^{n}(data_i - model_i)^2/n$).



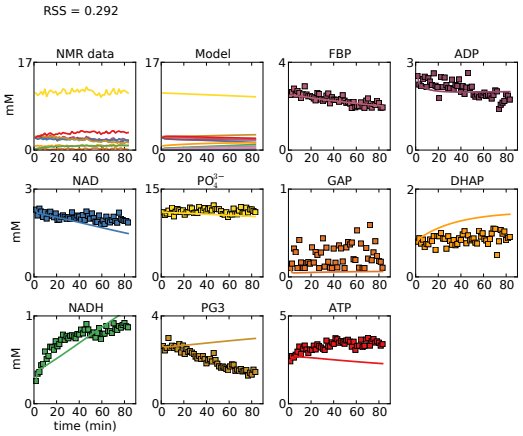
(e)



(f)



(g)



(h)

Figure 4.8: continued

determined by sonicating the experimental cells and performing a Bradford assay (as in Section 3.3). Maximal glucose uptake rate was 0.087 ± 0.009 µmol.min^{-1}.mg^{-1}; almost identical to [126] at a growth rate of 0.55 h^{-1}.

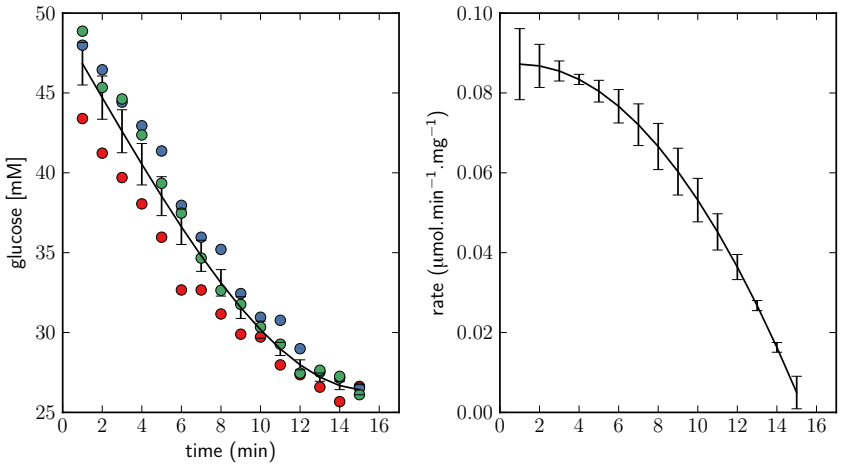


Figure 4.9: A ^{13}C glucose uptake assay. Left – Log-phase cells were incubated with 50 mM ^{13}C glucose and the reaction monitored on a 600 MHz NMR spectrometer. Concentration–time series were fitted with splines to extract a rate approximation at each data point. Right – Extracted rates from splines fitted to the concentration data. The assay was performed in triplicate and error bars indicate SEM. Black lines indicate the mean of three splines. Rates were normalised to protein concentration (determined using the Bradford assay, Section 3.3, mean total protein concentration was 8.2 mg/ml).

4.3.2 *In vivo* fermentation products

To assess the ratios of fermentation products, an *in vivo* ^{13}C glucose fermentation was performed in duplicate, and cell extracts were produced for NMR quantification. 2× 1 l microaerobic cultures of late-logarithmic phase cells (0.53 OD) were harvested by centrifugation for 7 min at 5,000 rpm (∼4200 × g) and a temperature of 4 °C. Pellets were resuspended in 50 mM PIPES buffer (pH 7.2) to 50 ml as a washing step and were pelleted again by centrifugation (4,500 rpm (∼5,600 × g) for 10 min at 4 °C). This pellet was resuspended in 4 ml 50 mM PIPES (pH 7.2). 2 incubations were performed in 50 mM PIPES pH 7.2, each containing 3.6 ml cells, and 0.4 ml 0.5 M $\frac{1}{2}$D-glucose-1-^{13}C:$\frac{1}{2}$D-glucose-^{12}C (to a final concentration 50 mM). OD_{600} was estimated by spectrophotometry of serially diluted cells and total protein was determined by the Bradford assay (as in Section 3.3). Mean OD_{600} was about 90, and mean protein concentration was 10.5 mg/ml. Fermentations were performed at 25 °C and 1 ml of culture was extracted from the larger culture and quenched at 3, 4.5 and 6 min by the addition of 250 µl of ice-cold 35 % perchloric acid. Extracts were centrifuged for 10 min at 13,000 rpm using a micro-centrifuge after which 1 ml was removed and neutralised with 100 µl 7 M KOH and centrifuged once more. About 0.6 ml of supernatant was salvaged. For NMR

quantification, 10% D_2O was added to the extracts. ^{13}C NMR was performed using a 300 MHz Varian spectrometer at 75.37 MHz with proton decoupling and NOE on, using a 90 ° pulse angle and a temperature of 25 °C. A repetition time of 16.3 s (1.3 s acquisition, 15 s relaxation) was used. To calibrate for possible incomplete relaxation and the NOE, a relaxed (25 s) spectrum with NOE excluded was acquired and integrals were adjusted accordingly. Peaks were normalised to total ^{13}C content for quantification. Rates were determined by fitting a linear equation to the means of the two experiments weighted by standard error using the LM algorithm [168] as implemented in scipy.optimize [163].

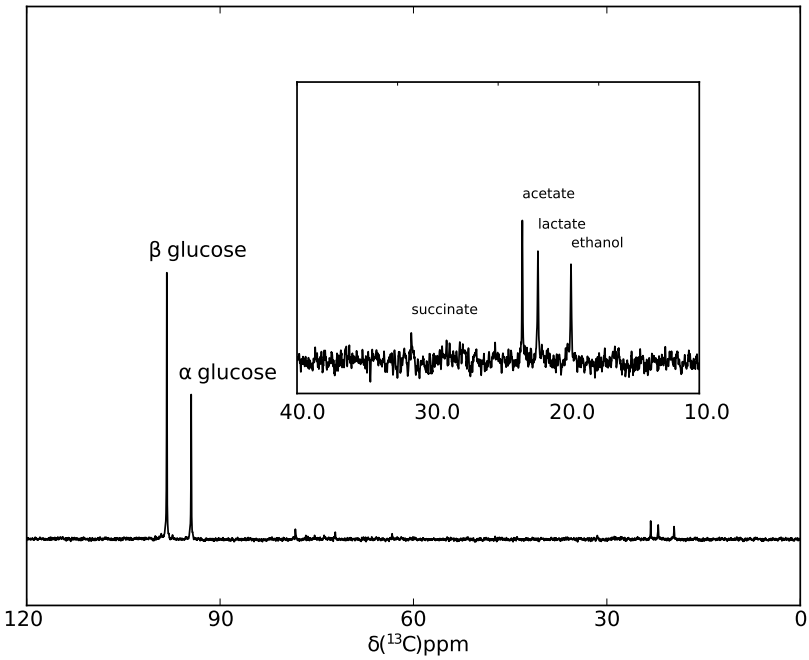


Figure 4.10: Fermentation products appear up field of glucose during a ^{13}C glucose fermentation assay. Extracts were made of an *in vivo* glucose fermentation assay at various time points to determine the ratio and flux distribution of fermentation products. This spectrum represents a 10 h NMR acquisition of a selected extract, on a 300 MHz spectrometer at 25 °C using a pulse angle of 90°, proton decoupling (without NOE) and a repetition time of 31.5 s (1.5 s acquisition, 30 s relaxation) to collect 1150 transients. The inset shows the fermentation products.

Fig. 4.10 shows a typical ^{13}C NMR spectrum produced during fermentation of glucose. Fig. 4.11a shows the quantified glucose and fermentation products in the three extracts and Fig. 4.11b shows the fitted flux values. Under the current experimental conditions, acetate production dominated followed by ethanol production. Significantly lower amounts of lactate and succinate were produced demonstrating that *in vivo*, the fermentation route through PFL is dominant. Fig. 4.11c shows a set of summarised published measurements of the mixed acid fermentation products under various conditions of oxygen availability. Key features of the published observations are the predominance

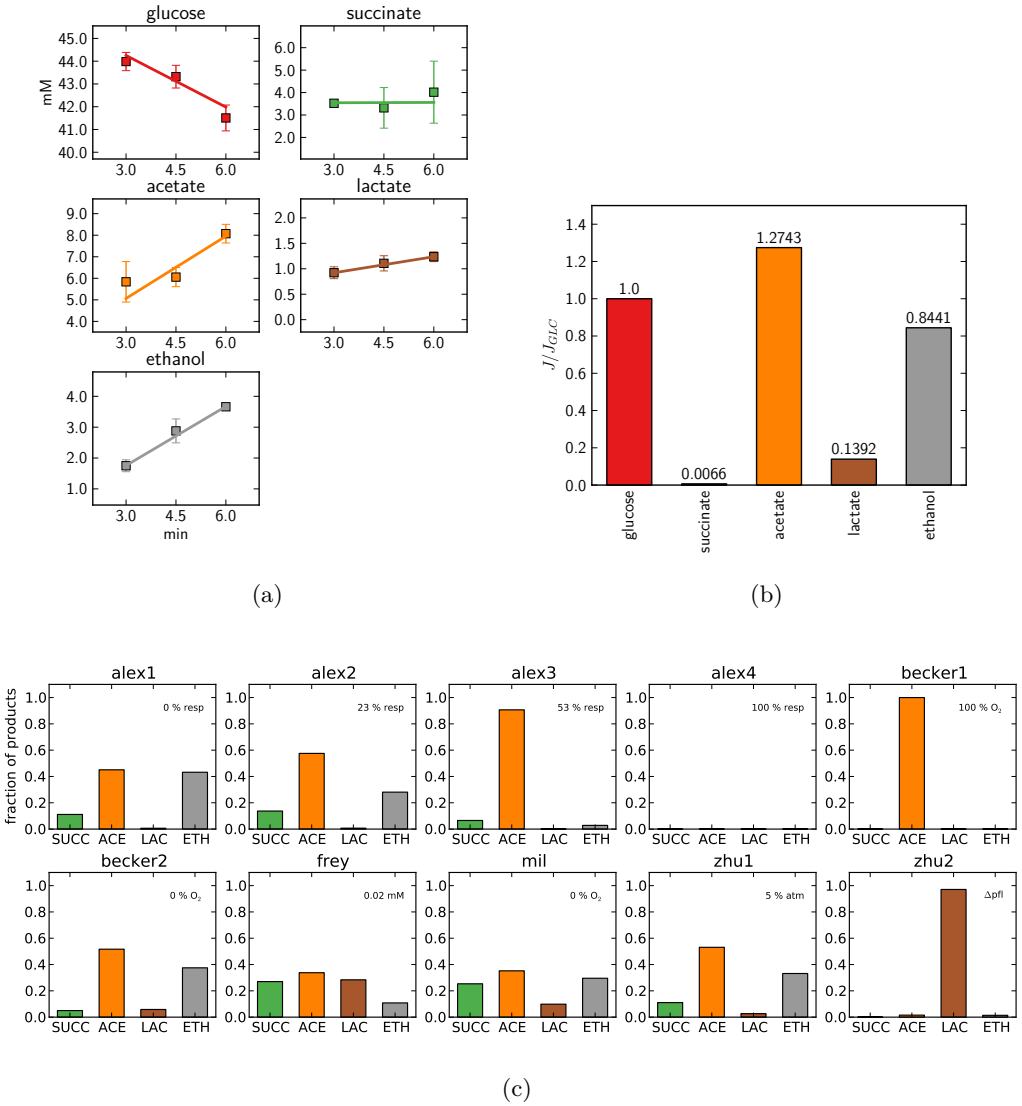
of acetate production and the increasing acetate to ethanol fermentation ratio with increasing O_2 (see Fig. 4.11c blocks "alex1" $\rightarrow$ "alex4" representing [132] for a salient example). As the culture conditions approach full anaerobiosis, acetate and ethanol production equalise. This is explained by the background NADH consumption by the high affinity $cytb_d$ that is up-regulated under microaerobic conditions [129] (surveyed in Section 1.6.3) which has been argued to "scavenge" residual O_2 under microaerobic conditions, protecting PFL from O_2 inactivation [132]. Consequently, a lower $\frac{[NADH]}{[NAD^+]}$ ratio alleviates the redox burden on the ethanol pathway, freeing the cell to redirect flux to acetate and produce additional ATP. The study by Zhu and Shimizu [260], represented in Fig. 4.11c (block "zhu2"), employed the same culturing technique as in this study to induce microaerobiosis by cultivating the cells with a gently spinning magnetic stirrer bar. The presence of background NADH consumption identified models O, P and Q as likely candidates. As such, model R is henceforth excluded.

4.3.3 Towards a steady state: influx and efflux routes

The steady state of a system, that is when input and output fluxes are equal and metabolite pools are unchanging ($\frac{\partial \mathbf{S}}{\partial t} = \mathbf{0}$, where $\mathbf{S}$ is a vector of metabolites), provides the platform for understanding the control properties of the system under study [14]. To acquire a steady state, the three putative models that were constructed (O, P and Q) required expansion to include the PTS glucose import system, the acetate, ethanol and succinate efflux routes, and steady state metabolite concentrations of fixed species (e.g. PO_4^{3-}, HCO_3^-). All kinetic parameters and rate equations are aggregated in Section 4.5.

The model of the PTS produced by Rohwer *et al.* [68] was included in the current models (summarised in Section 1.2.2). The glucose PTS is a group transfer pathway in which a phosphoryl moiety is sequentially passed along four proteins (EI, HPr, EIIA, EIICB) from PEP to external glucose (reviewed by Meadow *et al.* [72]). The final step involves the transport and phosphorylation of external glucose into the cell as G6P mediated by the membrane-bound EIICB. This activity is similar to the two-component signalling pathways [129]. Due to the nature of group transfer, the protein components are not acting strictly as catalysts as at the completion of any particular reaction the components involved either lose or gain a phosphoryl group, and thus do not remain unchanged. The PTS has however, often been characterised as a series of enzyme catalysed reactions with an abundance of K_m values reported (e.g. [264–267]). These parameters are nonetheless valuable and have been used to model this pathway in detail, capturing the group-transfer kinetic behaviour using elementary reactions. The PTS model thus consists of ten association/dissociation reactions, each with a forward and reverse rate constant. MCA had previously been performed experimentally *in vivo* [268], showing that the membrane bound EIICB dominated control of α-MG uptake. This is reflected in the control properties of the PTS model as the highest flux response coefficient belongs to the concentration of EIICB ($R^J_{EIICB} = 0.9$).

The three mixed acid fermentation routes already discussed (succinate, ethanol and acetate,



(a) (b)

(c)

Figure 4.11: The ratios of glucose fermentation products. (a) Glucose and fermentation products produced by log-phase cells were assayed in duplicate on a 300 MHz NMR spectrometer and fitted with a linear function to approximate (b) steady-state flux normalised glucose uptake. It is clear from the ratio of fermentation products and steady-state flux distribution that the route through PFL to acetate and ethanol dominates. (c) Literature fermentation product distributions normalised to total fermentation product. Titles of blocks refer to published articles, growth conditions are indicated in top-right corners: alex1 → alex4 [132] — % aerobiosis defined as % glucose converted to CO_2, or simply % respiration; becker1/2 [261] — conditions are saturating O_2 or anaerobic growth under N_2; frey [262] — 0.02 mM O_2 is equivalent to 0.39 mmHg or $0.5\times\ 10^{-3}$ atm; mil [263] — anaerobic growth; zhu1/ 2 [260] — microaerobic growth by slow agitation of cultures with a magnetic stirrer-bar (50 rpm), dissolved oxygen decreased to less than 5% of air saturation. Note that zhu2 demonstrates the effect of a $\Delta pflA$ mutant.

Fig. 4.3, LDH was already present during fitting of lower glycolysis) were added to the putative models to complete them. Where available, literature values were used for all of the parameters in these pathways. The ethanol and acetate fermentation pathways which convert acetyl-CoA to their respective fermentation products are both composed of two reactions and were each simplified by combining these two reactions into a single bi-substrate irreversible Michaelis-Menten equation (Section 4.5.2). The succinate pathway was modelled with mostly reversible equations parameterised with literature kinetic values for all reactions. An additional reaction (FRDII) was included to reoxidise the menaquinol pool produced by fumarate reductase (FRD). Additionally, the PEP carboxykinase (PCK) reaction was included as it has been observed during growth on glucose, producing an ATP-dissipating futile cycle [122, 150, 248]. As FUM exists as three isoenzymes in *E. coli*, the contribution of each isoenzyme to the rate in traditional kinetic assays is indeterminable. As such the intermediate isoenzyme's kinetic parameters were selected. Microaerobic maximal activities for all the succinate and ethanol/acetate enzymes were taken or calculated from [150] with the exception of FRD, and PFL which was very low, possibly due to oxygen inactivation which the authors do not mention taking any precautions against. PFL catalyses the splitting of pyruvate into the energy rich acetyl-CoA and formate, and acts as a "gate-keeper" of the lower mixed-acid fermentation reactions leading to acetate and ethanol, and competes with LDH for pyruvate [129]. PFL is very sensitive to the presence of oxygen, which reacts with a glycyl radical moiety in the peptide backbone inducing scission of the protein and inactivating the enzyme [269]. As an oxygen sensitive enzyme, PFL is functional only under anaerobic conditions and microaerobic conditions under which O_2 scavenging by the high affinity cytochrome b_d has been hypothesised to protect PFL from inactivation [132]. Additionally, PFL is converted into an inactive form under aerobic conditions and in stationary phase which protects the enzyme from denaturation [129]. This inactivation is performed by an ancillary enzyme, "enzyme 2" [270], which has previously been identified as alcohol dehydrogenase (ADH) [271] though this hypothesis has not gone unchallenged [272]. The maximal rate chosen reflects the fully active non-denatured state, which was determined experimentally by extracting and reactivating inactive PFL from anaerobic stationary phase cultures [270]. All rates were scaled to reflect an internal protein concentration of 250 $mg.ml^{-1}$ [273]. A non-specific mass-action ATP-consuming reaction ($v = k_{ATP}ATP$) was included in each model with the rate constant scaled to match the rate of ATP production in the simulated steady states. Additionally, the ADK reaction was removed when present to exclude AMP from the adenylate moiety-conservation relationship, and the $\frac{[ATP]}{[ADP]}$ ratio was represented as a single moiety RA with a fixed total moiety concentration.

Models were constructed with fixed external metabolites (glucose, acetate, ethanol, succinate, and lactate). The microaerobic $\frac{[ATP]}{[ADP]}$ ratio was set to $\frac{9.12mM}{6.24mM} = 1.46$ [262], and calculated assuming 2.5 µl/mg dry weight [274]. The microaerobic $\frac{[NADH]}{[NAD^+]}$ ratio was set to $\frac{0.46mM}{1.14mM} = 0.4$ [132], also calculated assuming 2.5 µl cytosolic volume/mg DW [274]. PO_4^{3-} was fixed at 20 mM [275], and due to the undetermined rate of diffusion of CO_2 out of the cell, it was assumed that HCO_3^-

would be present at saturating concentrations as most of the glycolytic flux tends to pass through PFL (Fig. 4.11) producing CO_2 via FHL, thus it was fixed at 10 mM. Total Acetyl-CoA/CoA was set to 2.0 mM [276]. $K_{eq}GAPDH$ was increased to 0.0056 [24] to reflect the increased ionic concentration *in vivo* [259]. It is not clear to what degree the $cytb_d$ NADH consumption would be achieved by NADH-dehydrogenase I (NDH1) or II (NDH2). In contrast to NDH2, NDH1 is able to produce an electrochemical gradient which could drive the synthesis of ATP [235], and NDH2 is down-regulated under low oxygen conditions. However, as the ATP consumption rate was scaled to achieve a steady state, including ATP as a product of respiration ($vNADH$) had no effect on the steady state analyses conducted in this study (assuming up to 4 H^+ translocated per NADH, indicating total NDH1 presence [235], and assuming 3.3 H^+ per ATP produced by ATP synthase [277]).

Upon initial simulation of the complete models, steady states were achieved, and all model glycolytic fluxes were around 14 $mM.min^{-1}$ using the wildtype PTS model parameters as reported by Rohwer *et al.* [68]. This value was just over half of the experimentally determined uptake rate of 21.75 ± 2.25 $mM.min^{-1}$ (Fig. 4.9). The steady state flux value achieved was essentially the same as that produced by the PTS model alone with fixed external metabolites [68]. The fact that the steady state flux is identical to that produced by the PTS model in isolation served as an initial indication that the PTS may be dominating glycolytic flux control. An initial examination of the flux response coefficients of the models indicated unilaterally that the parameter having by far the greatest effect on steady state flux was $\alpha_{FBP}PK$; the parameter inducing allosteric activation of PK by FBP. This resulted in a low PEP concentration, and thus, by virtue of the competition between the PTS and PK for PEP, a low steady state flux. $\alpha_{FBP}PK$ was reduced to a quarter of its original value in all models to reflect the fact that the *in vivo* environment (especially *in vivo* PO_4^{3-} concentrations) almost entirely abolishes the allosteric effect of FBP on PK [257]. All three models subsequently achieved the experimentally determined uptake rate of 21.75 ± 2.25 $mM.min^{-1}$. Thus, no changes to the published PTS model were necessary. Steady state data for the three models are shown in Fig. 4.12.

4.3.4 Comparison of steady state model behaviour with published observations of glycolysis in *E. coli*

All three models produced similar fermentation flux distributions to the experimentally determined values (Fig. 4.11b) with acetate flux dominating, and ethanol flux second to this; lactate and succinate fluxes were in the minority. This distribution is similar to several of the observed literature results (Fig. 4.11c, "alex2", "becker2", "zhu1"), though the lactate contribution in the models is lower. A cursory inspection of the steady state concentrations achieved by the models reveals a thermodynamic tendency for FBP and DHAP (models O and Q) and 3PG (all models) to accumulate, which is expected if the adjacent reactions are at or near equilibrium. A similar increase in FBP has been measured [276, 278, 279] and appears in at least one glycolytic modelling at-

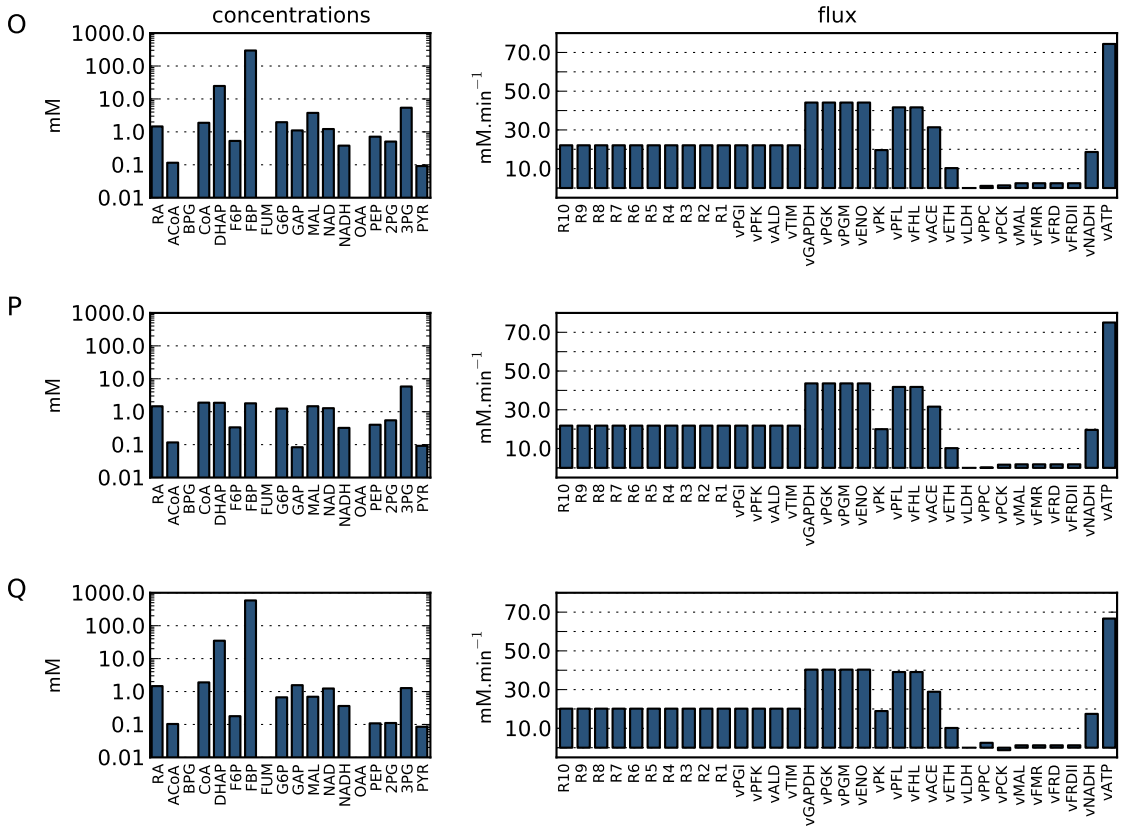


Figure 4.12: Steady state intracellular metabolite concentrations and fluxes for models O, P and Q. Models O, P and Q were expanded with literature parameter values to include the acetate, ethanol and succinate mixed acid fermentation pathways and the PTS glucose import pathway. The concentration of EIICB was altered to obtain the correct experimental glucose uptake flux. The steady state properties for each model are shown including concentrations (left) and fluxes (right). $vATP$ is a non-specific ATP-consuming reaction set to match the rate of ATP production.

tempt with *S. cerevisiae* [24]. Models O and Q produced significantly high and non-physiological steady state FBP concentrations of 296 mM, and 585 mM; model P maintained FBP at a lower, physiological concentration of 1.79 mM.

The models generally maintained the $\frac{[NADH]}{[NAD^+]}$ ratio close to the published value that the models were initialised with (O = 0.31 mM, P = 0.25 mM, Q = 0.29 mM) and remained within the literature range for microaerobiosis (0.1–0.4 mM [132, 260]). ATP and ADP concentrations were maintained at literature levels by the non-specific ATP consuming reaction. The steady state PEP concentration achieved by the models, which is key to modulating the PTS flux and the distribution of fermentation fluxes at the PEP/PYR node, was within the literature range (PEP: O = 0.71 mM, P = 0.11 mM, Q = 0.40 mM), while the PYR concentrations were somewhat lower than published values and all around 0.09 mM, possibly due to the choice of a fully active PFL at a rate determined under full anaerobiosis. The activity of PFL has been shown to vary up to five-fold over the microaerobic range of 0 %–40 % aerobiosis [132] and PFL is both irreversibly denatured by O_2 and reversibly inactivated by an ancillary enzyme under microaerobic conditions. The lower LDH flux in the models compared to experimental data may also be indicative of a high PFL rate in the models. Indeed, reducing the maximal PFL rate five-fold in model P, produced a significantly higher steady state PYR concentration of around 0.5 mM, and increased LDH activity significantly while leaving the PEP concentration and steady state flux largely unchanged. This relationship is explored further in Section 5.7. For the purpose of comparison, a table of selected literature values for intracellular metabolites is included below (Table 4.3).

It has been shown that correction of *in vitro* kinetic parameters to reflect *in vivo* conditions was able to ameliorate the tendency of FBP concentrations to blow up in a steady state model of glycolysis in yeast [258]. Inspection of the concentration control coefficients (Section 1.5 describes the quantification of state variable sensitivity towards a particular reaction rate) for the models with high FBP concentrations indicated a unanimous dependence of the steady state FBP concentration on the rate of $R8$ of the PTS, $vGAPDH$ and to a lesser extent reactions involving redox agents (e.g. $vNADH$ and $vETH$). It is likely that the positive effect of $R8$ on the FBP concentration is as a result of total flux control, that is, increasing the $R8$ rate increases pathway flux and many other metabolites too (this point will be revisited in Chapter 5). However, the $vGAPDH$ reaction exerts significant control only over the preceding metabolites FBP, DHAP and GAP, but most significantly toward FBP (C^{FBP}_{vGAPDH}: O = -3.33, P = -4.82, Q = -2.85). $vNADH$ exhibits a similar control pattern (C^{FBP}_{vNADH}: O = -0.92, P = -4.4, Q = -1.3). This suggests that in the models, and possibly under microaerobiosis, $vNADH$ hold the GAPDH reaction in tension by reducing the $\frac{[NADH]}{[NAD^+]}$ ratio, and activating the GAPDH reaction which acts as a bottleneck to upper glycolytic flux, causing FBP, DHAP and GAP to accumulate. GAPDH kinetics may behave as differently under *in vivo* conditions as the thermodynamics do [259]. In all of these models, the GAPDH reaction exerted almost no control over flux. Doubling the maximal forward rate of GAPDH in all models and re-simulating reduced the FBP concentration significantly, while leaving flux values unchanged:

O 296.00 mM → 43.23 mM

P 1.80 mM → 0.51 mM

Q 585.03 mM → 50.82 mM

Table 4.3: Published intracellular concentrations of glycolytic metabolites under different growth conditions of WT *E. coli*.

	Microaerobic[a]				Aerobic						
	[279]	[262]	[132]	[260]	[46]	[126] $\mu = 0.1$	[126] $\mu = 0.55$	[125]	[278]	[280]	[276]
HXP[b]											8.8
G6P	0.0196			0.02	3.48			1.76	1.25	0.8	
F6P					0.6			0.42	0.95		
FBP	1.8372			1.836	0.272	0.92	0.46	0.67	5.1	1.9	15
GAP					0.218			0.43			
DHAP								0.231		0.2	0.37
3PG						1.67	0.42				1.5
2PG								1.23			
PEP	0.128			0.128	2.67	0.88	0.06	1.04	0.15	0.09 - 0.3	0.18
PYR	3.284	4.88		3.284	2.67	1.64	0.28	1.71		0.39	
ATP	0.0348	9.12		1.224	4.27	0.94	1.2	3.54	3.5	2.6	9.6
ADP	0.152	6.24		0.152	0.595	0.32	0.63	0.33	0.15	0.82	0.56
$\frac{[ATP]}{[ADP]}$	0.23	1.46	8.05	7.18	2.94	0.19	10.73	23.33	3.17		17.14
NAD^+	0.0572		1.14	0.0572	1.47					1.26-2.08	2.6
NADH	0.0072		0.46	0.0072	0.1			0.135			0.83
$\frac{[NADH]}{[NAD^+]}$	0.13		0.40	0.13	0.07						0.32
aCoA	0.0264			0.028		1.42	0.68	0.145	0.45	0.35	0.61
CoA											1.4

[a] Calculated from dry weight assuming 2.5 µl/mg DW [274]

[b] total sugar phosphates including G6P, F6P and glucose 1-phosphate

A study of *pfl* knockout strains of *E. coli* grown under various conditions of O_2 availability has likewise demonstrated significant FBP (and PYR) accumulation [279] (Table 4.4). The simulated concentrations for the models in Fig. 4.12 (especially models O and Q) reproduce some of these key findings. The simulated FBP concentrations at steady state for models O, P and Q are mentioned above; Q producing a significantly higher FBP concentration than the others, model P resembling the wildtype concentration. By doubling the activity of GAPDH Models O and Q were able to reproduce a FBP concentration closer to the value measured in the Δpfl mutants

(FBP concentration was shown to increase as high as 66 mM in the *pflA* mutant during aerobic growth on pyruvate). Steady state PYR concentrations are shown to increase in the *pfl* mutants, a behaviour mimicked to a degree by the models. The $\frac{[ATP]}{[ADP]}$ ratios selected for the models are an order of magnitude higher than those reported for the pfl mutant study, however, the ATP and ADP concentrations reported by Zhu and Shimizu [279] are suspiciously low when compared to other literature values (ATP concentration is on average around 100 times lower when compared to the publications in Table 4.3). This may be due to mislabelling of the data tables in Zhu and Shimizu [260, 279]. In summary, model P seems most capable of reproducing the published degree of FBP accumulation. Model O and Q accumulate FBP to a far higher degree, however, the low PYR concentrations produced by the models suggest that if anything, PFL activity in these models is high (see above). Altering PFL activity has no effect on FBP and thus the bottleneck at GAPDH appears to be the cause of this accumulation.

Under low O_2 availabilities where ATP generation depends on substrate-level phosphorylation, concentrations of glycolytic enzymes are up regulated [150] to compensate for the lack of respiratory ATP production, and indeed an increase in the degree of saturation of some glycolytic enzymes may be noticed due to the higher burden on the glycolytic pathway, thus diminishing the control that ATP demand exerts over glycolysis. Under these conditions, and more so when PFL activity is diminished, FBP may accumulate as the early preparative reactions of glycolysis outpace the latter fermentative reactions.

Table 4.4: Intracellular concentrations of glycolytic metabolites for growth of *pfl* mutants of *E. coli* on glucose under different oxygen availabilities [279]. Concentrations are in mM and are calculated from dry weight assuming 2.5 µl/mg DW [274]

	wt	$\Delta pflB$	$\Delta pflA$	O	P	Q
G6P	0.0196	0.0532	0.4372	1.97	1.24	0.67
FBP	1.8372	16.528	6.564	296.0	1.80	585.03
PEP	0.128	0.336	0.048	0.71	0.40	0.11
PYR	3.284	27.188	10.196	0.09	0.09	0.09
aCoA	0.0264	0.028	0.02	0.11	0.12	0.10
ATP	0.0348	0.0176	0.0208			
ADP	0.152	0.1368	0.078			
AMP	0.0856	0.0772	0.0788			
NADH	0.0072	0.0116	0.0224			
NAD^+	0.0572	0.02	0.024			
ATP/ADP	0.23	0.13	0.27	1.46	1.46	1.46
$NADH/NAD^+$	0.13	0.58	0.93	0.31	0.25	0.29

4.3.5 *In vivo* observation of glycolytic intermediate accumulation

To assess the possibility of various patterns of intracellular accumulation of FBP, DHAP and/or 3PG as predicted by the models, fermenting cells were monitored *in vivo* using ^{31}P NMR. A specific concern was whether the cells used for *in vivo* NMR studies grown under microaerobic conditions are overwhelmed by high diffused O_2 concentrations during the preparative steps, causing PFL to become denatured and the cells to behave like *pfl* mutants. NMR has been successfully employed to quantify metabolite changes *in vivo* [19, 28, 66, 202, 281]; albeit with notoriously low sensitivity [18]. Due to the low sensitivity of this technique, very high cell densities are required to achieve a sufficient signal to noise ratio.

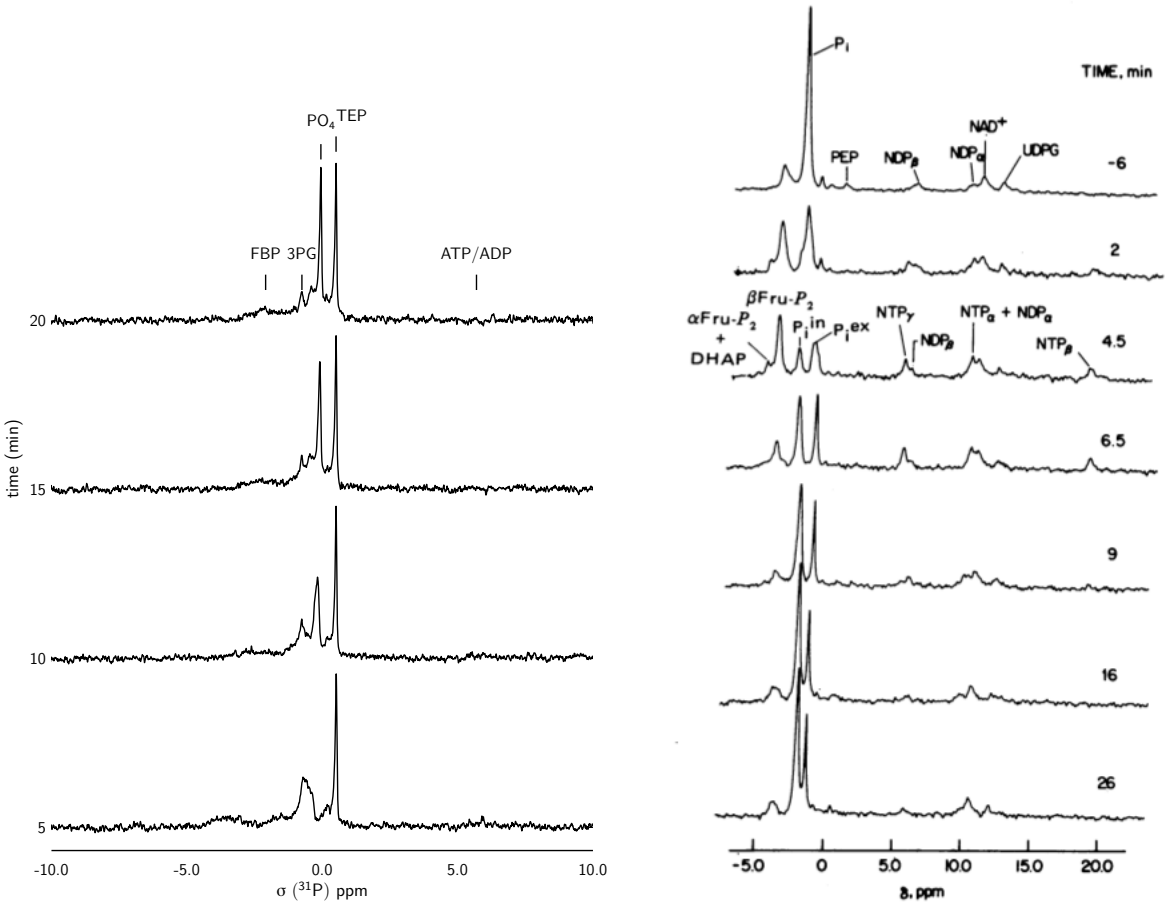
A 3 l culture of *E. coli* cells was grown in M9 medium under microaerobic conditions according to the protocol described in Section 3.3 and harvested in mid-log phase (OD 0.45) by centrifugation for 7 min at 5,000 rpm (∼4200 × g) and a temperature of 4 °C. Pellets were resuspended in 50 mM PIPES buffer (pH 7.2) to 50 ml as a washing step and were pelleted again by centrifugation (4,500 rpm (∼5,600 × g) for 10 min at 4 °C). This pellet was resuspended in a minimal amount (2 ml) 50mM PIPES (pH 7.2). For *in vivo* ^{31}P NMR, a 3 ml assay volume was used and a final cell density of ∼ 500 OD_{600}. The incubation was performed with 50 mM glucose, 5 mM TEP as an internal standard and 10 % D_2O in a 10 mM NMR tube.

^{31}P NMR was performed on a 600 MHz Varian spectrometer at 242.88 MHz, and a 90 ° pulse width with a repetition time of 0.5 s (0.5 s acquisition, 0.0 s relaxation) to collect 600 transients per FID (5 min per FID). A low power proton decoupling was used (including NOE) to avoid excessive heating of the sample. These parameters are similar to the protocol employed by [281] who used a 60 ° pulse width, 0.357 s repetition time and collected 800 transients per FID at a lower cell density of 180 OD_{600} and a lower power spectrometer (400 MHz). They did however use 15 mm NMR tubes. The incubation was started by the addition of glucose and monitored for 20 minutes (4 FIDs, Fig. 4.13a).

Though spectra were not calibrated for relaxation, it is qualitatively clear that 3PG accumulates *in vivo* and likely FBP too, though not necessarily to the concentrations observed for the *pfl* mutants in [279]. It is likely that PFL was partially deactivated under the *in vivo* experimental set-up, though the degree of deactivation is unclear. The accumulation of FBP has been previously observed anaerobically under similar experimental conditions [202] (Fig. 4.13b) and is thus probably a general phenomenon of the over-burdening of the glycolytic pathway during fermentative growth due to the cell's reliance on substrate-level phosphorylation for ATP generation.

4.4 Conclusion

A variation on the new methodology involving NMR presented in Chapter 3 has been developed and applied to the modelling of the complex kinetics of a cell extract in microaerobically fermenting *E. coli* cells. ^{31}P NMR time courses were fitted with an appropriate model of the lower half of glycolysis using a genetic algorithm to approximate kinetic parameters. Genetic algorithms proved to be the most effective metaheuristic algorithm by virtue of their ability to search parameter space widely without sacrificing the ability to identify a particular minimum solution. Ideally these parameters can be fine-tuned using a fitting algorithm that performs well locally. The Levenberg-Marquardt algorithm is particularly useful as it provides an error estimate on the fitted parameters (if the model is constrained by the data). After the addition to the lower-glycolytic model of the upper-glycolytic reactions developed and parameterised in Chapter 3, a selection of three models (O, P and Q) was made based on the Akaike information criterion and the ability of the models to emulate *in situ* ^{13}C NMR experimental data. After expanding these models to include the PTS and fermentative pathways, the steady state properties were explored. The models were scrutinised by comparing their fermentative behaviour with data from ^{13}C NMR quantification of the fermentation balance in quenched cell extracts, and ^{31}P NMR of *in vivo* accumulation of intracellular glycolytic intermediates. Model behaviour was also compared to published intracellular concentrations and fluxes, as well as fermentation outputs. The models were generally able to reproduce the expected fermentative flux distribution, and presented various degrees of success at predicting intracellular metabolite concentrations. Interestingly, two of the models predicted a large accumulation of FBP, DHAP and 3PG which was observed in published studies and in *in vivo* NMR assays. Unlike



(a) (b)

Figure 4.13: ^{31}P NMR analysis of *in vivo* glycolytic intermediates. (a) A high density cell suspension was incubated with 50 mM glucose and monitored using ^{31}P NMR. Accumulation of intracellular 3PG is noticeable, and accumulation of FBP is very likely (especially indicated by the broad peaks area in the lowest block). (b) A similar ^{31}P NMR experiment was performed under anaerobic conditions [202] (image reproduced from cited reference). Spectra were collected at 145.7 MHz with a 90 ° pulse angle, a repetition time of 0.6 s. 200 transients were collected per FID. The buffer used was comprised of 20 mM phosphate/85 mM NaCl/50 mM MES/100 mM PIPES at pH 7.3. Incubation was started by the addition of 25 mM glucose. Note that in (a) the time axis proceeds upwards, and in (b) downwards.

the literature however, the accumulation of FBP in the models was not necessarily due to PFL inactivation but may be the result of a flux bottleneck at GAPDH as the earlier stages of glycolysis outpaced the latter (or likely a combination of both of these effects). To some degree this may be due to a slower rate of NADH reoxidation during fermentation, leading to a decrease in the GAPDH substrate NAD^+. By virtue of the ability of model P to emulate *in situ* data and the literature steady state fluxes and metabolite concentration, it was demonstrated to be superior to models O and Q.

4.5 Literature parameters and rate equations

4.5.1 Literature parameters: glycolytic enzymes

The following reactions were included in the full model of glycolysis. Literature parameter values that were employed are indicated in respective tables. Kinetic equations are presented in Table 4.5.

The PEP:phosphotransferase system (PTS)

GLC + PEP → G6P + PYR

R1: $\mathsf{PEP} + \mathsf{I} \leftrightarrow \mathsf{PYR\_P\_I}$

$k_1 * PEP * I - k_{-1} * PYR\_P\_I$

R2: $\mathsf{PEP\_P\_I} \leftrightarrow \mathsf{PYR} + \mathsf{IP}$

$k_2 * PYR\_P\_I - k_{-2} * PYR * IP$

R3: $\mathsf{IP} + \mathsf{HPr} \leftrightarrow \mathsf{I\_P\_HPr}$

$k_3 * IP * HPr - k_{-3} * I\_P\_HPr$

R4: $\mathsf{I\_P\_HPr} \leftrightarrow \mathsf{I} + \mathsf{HPrP}$

$k_4 * I\_P\_HPr - k_{-4} * HPrP * I$

R5: $\mathsf{HPrP} + \mathsf{IIA} \leftrightarrow \mathsf{HPr\_P\_IIA}$

$k_5 * HPrP * IIA - k_{-5} * HPr\_P\_IIA$

R6: $\mathsf{HPr\_P\_IIA} \leftrightarrow \mathsf{HPr} + \mathsf{IIAP}$

$k_6 * HPr\_P\_IIA - k_{-6} * HPr * IIAP$

R7: $\mathsf{IIAP} + \mathsf{IICB} \leftrightarrow \mathsf{IIA\_P\_IICB}$

$k_7 * IIAP * IICB - k_{-7} * IIA\_P\_IICB$

R8: $\mathsf{IIA\_P\_IICB} \leftrightarrow \mathsf{IIA} + \mathsf{IICBP}$

$k_8 * IIA\_P\_IICB - k_{-8} * IIA * IICBP$

R9: $\mathsf{IICBP} + \mathsf{GLC} \leftrightarrow \mathsf{IICB\_P\_GLC}$

$k_9 * IICBP * GLC - k_{-9} * IICB\_P\_GLC$

R10: $\mathsf{IICB\_P\_GLC} \leftrightarrow \mathsf{IICB} + \mathsf{G6P}$

$k_{10} * IICB\_P\_GLC - k_{-10} * IICB * G6P$

The primary glucose import pathway in *E. coli* is the phosphotransferase system (PTS) which is a group-transfer pathway, sequentially transferring a phosphate moiety along a series of enzymes (EI, HPr, EIIA, EIICB) from PEP to external glucose, which is imported and phosphorylated (reviewed by Meadow *et al.* [72]). This pathway has been modelled in detail to capture the group-transfer kinetic behaviour using elementary reactions [68]. This was achieved by separating the five phosphotransfer reactions into ten elementary reactions of association and dissociation represented above (R1 – R10, P represents a phosphate moiety and complexes are conjoined with underscores). Rate constants were scaled to the same units as the model:

Param.		Ref.
k_1	$1,960 \times 10^3$ mM^{-1}min^{-1}	[68]
k_{-1}	$480,000$ min^{-1}	”
k_2	$108,000$ min^{-1}	”
k_{-2}	294×10^3 mM^{-1}min^{-1}	”
k_3	$14,000 \times 10^3$ mM^{-1}min^{-1}	”
k_{-3}	$14,000$ min^{-1}	”
k_4	$84,000$ min^{-1}	”
k_{-4}	$3,360 \times 10^3$ mM^{-1}min^{-1}	”
k_5	$21,960 \times 10^3$ mM^{-1}min^{-1}	”
k_{-5}	$21,960$ min^{-1}	”
k_6	$4,392$ min^{-1}	”
k_{-6}	$3,384 \times 10^3$ mM^{-1}min^{-1}	”
k_7	880×10^3 mM^{-1}min^{-1}	”
k_{-7}	880 min^{-1}	”
k_8	$2,640$ min^{-1}	”
k_{-8}	960×10^3 mM^{-1}min^{-1}	”
k_9	260×10^3 mM^{-1}min^{-1}	”
k_{-9}	389 min^{-1}	”
k_{10}	$4,800$ min^{-1}	”
k_{-10}	0.0054×10^3 mM^{-1}min^{-1}	”

Fructose 1,6-bisphosphate aldolase (ALD)

$$\text{FBP} \leftrightarrow \text{DHAP} + \text{GAP}$$

E. coli has two FBP aldolases which split FBP into DHAP and GAP: a class I enzyme (*fbaB*) that is only present during gluconeogenesis (e.g. growth on 3-carbon substrates) [282]; and a class

II metal-dependent enzyme present during glycolysis (*fbaA*). The class II enzyme has a very high optimal pH of 9 [283] and requires a divalent metal cation for activity, ideally Zn^{2+}, though other metals are able to induce catalysis albeit at significantly lower rates [284]. The activity of aldolase increases up to 2-fold in microaerobic and anaerobic cultures [285]. Under standard conditions the reverse reaction is favoured about 10-fold [286]. ALD has been shown to be mildly inhibited by PEP [222].

A uni-bi reversible Hill equation was used to model ALD with the following parameter values values taken from literature:

Param.		Ref.	
V_f	3.3 µmol/min/mg	[283]	Corrected for lower than optimal pH as described in the reference
$FBP_{0.5}$	0.29 mM	[283]	Corrected for lower than optimal pH as described in the reference
$DHAP_{0.5} = GAP_{0.5}$	0.13 mM	[287]	Values for DHAP and GAP product binding are set to equal as in [115, 288]
K_{eq}	0.1 mM	[288]	This value is dependent upon Mg^{2+} concentrations and may extend to as high as 0.17 mM [286] at intracellular concentrations (2–3 mM Mg^{2+} [289])

Triosephosphate isomerase (TPI)

DHAP ↔ GAP

The dimeric TPI of *E. coli* catalyses the high rate interconversion of GAP and DHAP. Under standard conditions the reverse reaction is favoured 20-fold [290].

A uni-uni reversible Hill equation was used to model TPI with the following parameter values values taken from literature:

Param.		Ref.	
$DHAP_{0.5}$	2.8 mM	[283]	
$GAP_{0.5}$	0.3 mM	[283]	
K_{eq}	0.045	[290]	This value was determined at 38 °C. Correcting to reflect the current experimental temperature of 25 °C using the Van t'Hoff equation [291] yields a value of 0.047 (ΔH = -2.73 [292]); a negligible difference.

Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)

$$\mathrm{GAP} + \mathrm{PO_4^{3-}} + \mathrm{NAD^+} \leftrightarrow \mathrm{BPG} + \mathrm{NADH}$$

E. coli has two GAPDH enzymes GAPDH-A and GAPDH-B; the former is required for glycolysis [293]. GAPDH catalyses the conversion of GAP and orthophosphate into BPG producing a reducing equivalent (NADH). The GAPDH reaction is favoured in the reverse and the K_{eq} for GAPDH has been shown to be heavily dependent on ion concentration [259]; increasing 7-fold over an ion concentration range of 0.0–0.6 M.

A ter-bi reversible Hill equation was used to model GAPDH with the following parameter values values taken from literature:

Param.		Ref.	
$GAP_{0.5}$	1.1 mM	[294]	
$NAD^+_{0.5}$	0.9 mM	[294]	
${PO_4^{3-}}_{0.5}$	0.53 mM	[295]	
$BPG_{0.5}$	0.08 mM	[294]	
$NADH_{0.5}$	0.05 mM	[294]	
K_{eq}	$4 \times 10^{-4}\mathrm{mM}^{-1}$	[259]	This value was calculated to represent an *in vitro* ion concentration of 110 mM according to the data in the reference. For the *in vivo* models, a higher value of 0.0056 [24, 296] was used to reflect the higher intracellular ion strength.

Phosphoglycerate kinase (PGK)

$$BPG + ADP \leftrightarrow 3PG + ATP$$

The second kinase of glycolysis in *E. coli* catalyses the production of a high energy ATP molecule by extracting a PO_4^{3-} group from the energy-rich acyl-phosphate BGP. The *pgk* gene is up regulated in stationary phase and under anaerobic conditions [285, 297]. PGK requires Mg^{2+} for catalysis. Typical of kinases, PGK has a very high K_{eq} value [259] which is dependent upon Mg^{2+} concentration, changing from 700 to 4000 over the Mg^{2+} concentration range of 0–1.4 mM.

A bi-bi reversible Hill equation was used to model PGK with the following parameter values values taken from literature:

Param.		Ref.	
$ATP_{0.5}$	0.24 mM	[298]	
K_{eq}	4000	[259]	This value was calculated to represent an experimental *in vitro* Mg^{2+} concentration of 5 mM according to the data in the reference and assumed to be saturated both *in vitro* and *in vivo*.

Phosphoglycerate mutase (PGM)

$$3PG \leftrightarrow 2PG$$

E. coli has two PGM isoenzymes which interconvert 3PG and 2PG, one isoenzyme is 2,3-BPG-dependent (common to vertebrates), the other is co-factor independent but Mn^{2+}-dependent (common to higher plants) [299]. As 2,3-BPG was not present in experiments, the *in vitro* model represents only the co-factor independent enzyme which has a lower activity [300]. The reaction is favoured 10-fold in reverse and the K_{eq} is dependent upon pH [301].

A uni-uni reversible Hill equation was used to model PGM with the following parameter values taken from literature:

Param.		Ref.	
$3PG_{0.5}$	0.2 mM	[299]	
$GAP_{0.5}$	0.097 mM	[299]	
K_{eq}	0.1	[301]	This value was calculated to represent an *in vitro* pH of 7.2 according to the reference.

Enolase (ENO)

2PG ↔ PEP

ENO interconverts 2PG and PEP and is structurally dependent upon a divalent metal ion (e.g. Mg^{2+}, Zn^{2+}, Mn^{2+}) [302, 303]. The reaction is favoured in the forward direction [302] and exhibits a pH optimum of 8.1 [302, 303].

A uni-uni reversible Hill equation was used to model ENO with the following parameter values taken from literature:

Param.		Ref.	
$2PG_{0.5}$	0.1 mM	[303]	
K_{eq}	4.5	[302]	This value was calculated to represent an *in vitro* pH of 7.2 and a temperature of 298.15 K according to the reference.

Pyruvate kinase (PK)

PEP + ADP ↔ PYR + ATP

E. coli expresses 2 PK enzymes which produce ATP from the dephosphorylation of PEP: PK I (*pykf*), which is activated by FBP, inhibited by succinyl-CoA [304] and both activated and inhibited under different conditions by ATP [224], and is more stable than PK II, which is activated by AMP and unaffected by FBP [221]. Both enzymes are dependent on Mg^{2+} and show a cooperative binding toward the substrate PEP and Mg^{2+} which is abolished in PKI by FBP [224]. PKI predominates in glycolysis, PKII in gluconeogenesis [224]. The forward reaction is thermodynamically favoured [305].

A bi-substrate reversible Hill equation with product inhibition and FBP as an allosteric modifier was used to model PK with the following parameter values values taken from literature:

Param.		Ref.	
$PEP_{0.5}$	3.63 mM	[306]	
$ADP_{0.5}$	0.3 mM	[306]	
$FBP_{0.5}$	0.045 mM	[306]	
K_{eq}	3.89×10^4	[305]	
α_{FBP}	6.762		As FBP affects the affinity of PK for PEP [221], the parameter value was selected to reduce the PEP half-saturation constant from 3.63 to 0.08 in the presence of 1 mM FBP according to the values in Valentini *et al.* [306].

Phosphoenolpyruvate synthetase (PPS)

$$\mathrm{PYR} + \mathrm{ATP} \leftrightarrow \mathrm{PEP} + \mathrm{AMP} + \mathrm{PO}_4^{3-}$$

During gluconeogenic growth PPS allows *E. coli* to produce PEP from PYR at the expense of ATP; the enzyme is induced on the approach to steady state and acetate accumulation [134, 234].

A reversible bi-ter Hill equation was used to model PPS with the following parameters taken from literature:

Param.		Ref.
$PYR_{0.5}$	0.083 mM	[307]
$ATP_{0.5}$	0.028 mM	[307]
$PEP_{0.5}$	17.0 mM	[308]
${PO_4^{3-}}_{0.5}$	10.5 mM	[307]
K_{eq}	52 mM	[309]

Adenylate kinase (ADK)

$$2\mathrm{ADP} \leftrightarrow \mathrm{ATP} + \mathrm{AMP}$$

Adenylate kinase equilibrates the ratios of adenylates in the cell.

A reversible bi-bi Hill equation was used to model ADK with the following parameters taken from literature:

Param.		Ref.
$ATP_{0.5}$	0.048 mM	[310]
$AMP_{0.5}$	0.038 mM	[310]
$ADP_{0.5}$	0.85 mM	[310]
K_{eq}	1.4	[311]

Background NADH consumption

$$\text{NADH} \rightarrow \text{NAD}^+$$

NADH consumption was fitted with an irreversible Michaelis-Menten equation.

4.5.2 Literature parameters: mixed acid fermentation enzymes

The following enzymes were included in the sub-model of mixed acid fermentation. Literature values that were used are indicated in respective tables:

Lactate dehydrogenase (LDH)

$$\text{PYR} + \text{NADH} \leftrightarrow \text{LAC} + \text{NAD}^+$$

E. coli expresses three LDH enzymes which provide a means of regenerating oxidising intermediates for glycolysis under fermentative conditions [129]. Two of the enzymes are membrane-associated flavoproteins, associated with aerobic growth, and are thermodynamically constrained to oxidise lactate to pyruvate during respiration [129, 312]. The third LDH is a soluble, NADH-dependent D-lactate dehydrogenase (nLDH) which is up regulated under low pH and low oxygen conditions [313]. The nLDH reaction is favoured in the direction of pyruvate reduction (K_{eq} = 7.35×10^4) [314] and binds pyruvate cooperatively [237]. NADH-binding affected by the pyruvate concentration; $NADH_{0.5}$ ranges from 0.24 at 3 mM PYR to 0.56 mM at 15 mM PYR [237]. The enzyme is sensitive to pH (optimal at 7.5), and at pH levels below neutral displays substrate inhibition by pyruvate [237]. nLDH is vulnerable to auto-oxidative inactivation; this is ameliorated to a degree by the presence of NADH and NAD^+ [236].

A bi-substrate irreversible Hill equation was used to model nLDH with the following parameter values values taken from literature:

Param.		Ref.	
V_f	1.22 µmol/min/mg	[150]	This value was determined under microaerobic conditions. The reported value under aerobic conditions was 1.1 µmol/min/mg [237]
$NADH_{0.5}$	0.25 mM	estimated from [237]	

Pyruvate formate-lyase (PFL)

PYR + CoA ↔ Acetyl-CoA + Formate

PFL catalyses the splitting of pyruvate into the energy rich acetyl-CoA and formate, and acts as a "gate-keeper" of the lower mixed-acid fermentation reactions [129]. PFL is very sensitive to the presence of oxygen, which reacts with a glycyl radical moiety in the peptide backbone inducing scission of the protein and inactivating the enzyme [269]. As an oxygen sensitive enzyme, PFL is functional only under anaerobic conditions and microaerobic conditions under which O_2 scavenging by the high affinity cytochrome b_d has been hypothesised to protect PFL from inactivation [132]. PFL is inactivated under aerobic conditions and in stationary phase by an ancillary enzyme, "enzyme 2" [270]. The parameters chosen reflect the fully active non-denatured state.

A bi-substrate reversible Michaelis-Menten equation was used to model PFL with the following parameter values values taken from literature (note that during parameter fitting of the lower glycolytic model to NMR data in Section 4.2.4, an irreversible Michaelis-Menten equation was employed for PFL) :

Param.		Ref.	
V_f	4.125 µmol/min/mg	[270]	This rate reflects the total reactivated PFL (see text for details) and is adjusted to the experimental temperature of 25° according to the Q_{10} rule [291].
$PYR_{0.5}$	2.0 mM	[315]	
$CoA_{0.5}$	0.0068 mM	[315]	
$FORM_{0.5}$	24.0 mM	[315]	
$CoA_{0.5}$	0.051 mM	[315]	
K_{eq}	800	[315]	

Phosphate acetyltransferase + acetate kinase (ACE)

$$\text{Acetyl-CoA} + PO_4^{3-} + \text{ADP} \leftrightarrow \text{ACE} + \text{ATP} + \text{CoA}$$

The phosphate acetyltransferase (PTA) and acetate kinase (ACK) enzymes of *E. coli* respectively catalyse the phosphorylation of acetyl-CoA to acetylphosphate, and subsequently the transfer of the high energy PO_4^{3-} moiety to ATP [129]. ACK has a requirement for Mg^{2+} and exhibit a lability that is remediated by MgATP and acetate [316]; PTA is unstable but protected by sulphate [317].

The PTA and ACK reactions were combined into a single irreversible Michaelis-Menten rate equation.

Param.		Ref.	
V_f	0.062 µmol/min/mg	[150]	This rate was determined under microaerobiosis.
$aCoA_{0.5}$	0.33 mM	[318]	
${PO_4^{3-}}_{0.5}$	1.5 mM	[318]	
$ADP_{0.5}$	0.5 mM	[316]	

Acetaldehyde dehydrogenase + alcohol dehydrogenase (ETH)

$$\text{Acetyl-CoA} + 2\text{NADH} \leftrightarrow \text{ETH} + 2\text{NAD}^+ + \text{CoA}$$

The ethanol dehydrogenase fermentation route reoxidises two NADH species by a two-step process; both steps are catalysed by the same enzyme [129].

The two ethanol dehydrogenase reactions were combined into a single irreversible Michaelis-Menten rate equation.

Param.		Ref.	
V_f	0.75 µmol/min/mg	[150]	This rate was determined under microaerobiosis.
$aCoA_{0.5}$	0.007 mM	[319]	
$NADH_{0.5}$	0.025 mM	[319]	

Succinate fermentation pathway

$$\mathrm{PEP} + \mathrm{CO_3^{2-}} \leftrightarrow \mathrm{OXA} + \mathrm{PO_4^{3-}}$$
$$\mathrm{OXA} + \mathrm{NADH} \leftrightarrow \mathrm{MAL} + \mathrm{NAD^+}$$
$$\mathrm{MAL} \leftrightarrow \mathrm{FUM} + \mathrm{H_2O}$$
$$\mathrm{FUM} + \mathrm{MQOH} \leftrightarrow \mathrm{SUCC} + \mathrm{MQO^+}$$
$$\mathrm{MQO} + \mathrm{NADH} \leftrightarrow \mathrm{MQOH} + \mathrm{NAD^+}$$

The succinate fermentation pathway consists of Phosphoenolpyruvate Carboxylase (PPC), Malate Dehydrogenase (MDH), three Fumarase isoenzymes (FMR) and Fumarate Reductase (FRD), and oxidises 2 reducing equivalents under anaerobic conditions [129]; the second oxidation is indirect and takes place via the quinol pool in the plasma membrane [320]. As a convenience, the final oxidation step involving the reduction of a menaquinone (MQO) to a menaqiunol (MQOH) has been modelled with a simple mass-action equation; a high rate constant (2500 $\mathrm{mM^{-1}min^{-1}}$) ensures that menaquinone is reduced. The three FMR isoenzymes exhibit different binding kinetics, and as such the contribution of each isoenzyme in activity assays with crude cell extracts is indistinguishable. Thus the kinetics for the isoenzyme with the middle value for the binding constant of malate was selected (i.e. $K_m = 1.1$ mM). PPC was modelled as a reversible bi-substrate reaction with cooperative binding of the substrates [321]. Additionally, PPC has been shown to be inhibited by the Krebs cycle intermediate malate, and activated by aCoA [321]. These allosteric relationships have been included in the rate equation as V-type effects, using γ values (a parameter similar to the α allosteric modulation parameter which enable K-type effects) that match the inhibition and activation data presented by [321] (PEP concentration was assumed to be 2 mM).

During the parameter fitting of the lower glycolytic model to NMR data (Section 4.2.4), the equations for PPC, MDH and FMR were replaced with irreversible forms to reduce the number of parameters to be fitted. The reversible forms listed below are included in the final model. The individual reactions were modelled as follows:

Param.		Ref.	
PPC			Reversible bi-substrate Hill equation with allosteric inhibition by malate and activation by aCoA.
V_f	0.0975 µmol/min/mg	[150]	This rate was determined under microaerobiosis, corrected for temperature.
$PEP_{0.5}$	0.19 mM	[243]	
${CO_3^{2-}}_{0.5}$	0.1 mM	[243]	
$OXA_{0.5}$	2.1 mM	[322]	This parameter was determined for *Corynebacterium glutamicum*
${PO_4^{3-}}_{0.5}$	1.0 mM	-	This parameter was unavailable and thus estimated.
K_{eq}	6.6×10^5 mM	[323]	
h	2.5	[321]	

Param.		Ref.	
MDH			Reversible bi-substrate Michaelis-Menten equation
V_f	0.3825 µmol/min/mg	[150]	This rate was determined under microaerobiosis in reverse, the forward rate was calculated from the ratio of forward and reverse rates determined by Courtright and Henning [238]
$OXA_{0.5}$	0.049 mM	[324]	
$NADH_{0.5}$	0.061 mM	[324]	
$MAL_{0.5}$	2.6 mM	[324]	
$NADH_{0.5}$	0.026 mM	[324]	
K_{eq}	$34,965$	[325]	

Param.		Ref.	
FMR			Reversible bi-substrate Hill equation, three isoenzymes with distinct kinetics
V_f	0.03 µmol/min/mg	[150]	This rate was determined under microaerobiosis.
$MAL_{0.5}$	1.1, 0.63, 2.94 mM	[326]	
$FUM_{0.5}$	0.15, 0.39, 1.7 mM	[326]	
K_{eq}	0.24	[327]	
FRD			Irreversible bi-substrate Michaelis-Menten equation
V_f	5.3 µmol/min/mg	[328]	This rate was determined under anaerobiosis.
$FUM_{0.5}$	0.02 mM	[326]	
$MQOH_{0.5}$	0.004 mM	[326]	
FRDII			Mass-action equation
k_f	2500 $mM^{-1}min^{-1}$		This parameter was set to fully reduce MQO produced by FRD.

Phosphoenolpyruvate carboxykinase (PCK)

$$OXA + ATP \leftrightarrow PEP + ADP + CO_3^{2-}$$

PCK catalyses the regeneration of PEP at the expense of ATP as the first step of gluconeogenesis [329] and has been implicated in a futile cycle in *E. coli* [247] by virtue of the reaction's reversibility. A reversible bi-substrate Michaelis-Menten equation was used to model PCK. As HCO_3^- was fixed at 10 mM in the model, it was excluded from the rate equation and the equilibrium constant was corrected for this. PCK was modelled with the following parameters taken from literature:

Param.		Ref.	
V_f	0.03 µmol/min/mg	[150]	This rate was determined under microaerobiosis, corrected for temperature.
$OXA_{0.5}$	0.31 mM	[330]	
$ATP_{0.5}$	0.74 mM	[330]	
$PEP_{0.5}$	0.07 mM	[331]	
$ADP_{0.5}$	0.05 mM	[331]	
K_{eq}	0.0174	[332]	corrected to exclude HCO_3^-

4.5.3 Rate equations

The rate equations employed in the construction of the model are shown in Table 4.5. For details with regard to which equation was selected for a particular reaction see the preceding sections.

Table 4.5: Kinetic rate equations used in the lower glycolytic and fermentative sub-model. All multisubstrate Hill equations taken from [79]. s, p and m denote substrates, products and allosteric modifiers scaled by their half-saturation constants (i.e. $s = \frac{S}{S_{0.5}}$); V_f is the maximal forward rate; K_m is the Michealis constant; h is the Hill cooperativity coefficient; Γ is a mass-action ratio (i.e. $\Gamma = \frac{\Pi[products]}{\Pi[substrates]}$); K_{eq} is the equilibrium constant; α and γ are K- and V-type allosteric modifier terms which indicate inhibition ($\alpha < 1$) or activation ($\alpha > 1$).

uni-substrate irreversible Michaelis-Menten [80]	$v = V_f \frac{S}{Km+S}$
uni-substrate irreversible Hill [86]	$v = V_f \frac{S^h}{S_{0.5}^h+S^h}$
bi-substrate irreversible Hill [88]	$v = V_f \frac{s_1 s_2}{1+s_1^h+s_2^h+s_1^h s_2^h}$
uni-uni reversible Michaelis-Menten [83]	$v = V_f \frac{s(1-\frac{\Gamma}{K_{eq}})}{1+s+p}$
uni-uni reversible Hill [83]	$v = V_f \frac{s(1-\frac{\Gamma}{K_{eq}})(s+p)^{h-1}}{1+(s+p)^h}$
uni-uni reversible Hill with modifier [83]	$v = V_f \frac{s(1-\frac{\Gamma}{K_{eq}})(s+p)^{h-1}}{\frac{1+m^h}{1+\alpha m^h}+(s+p)^h}$
uni-bi reversible Hill [79]	$v = V_f \frac{s(1-\frac{\Gamma}{K_{eq}})(s+p_1p_2)^{h-1}}{1+p_1^h+p_2^h+(s+p_1p_2)^h}$
bi-bi reversible Hill[79]	$v = V_f s_1 s_2 \left(1-\frac{\Gamma}{K_{eq}}\right)\left(\frac{(s_1+p_1)^{h-1}}{1+(s_1+p_1)^h}\right)\left(\frac{(s_2+p_2)^{h-1}}{1+(s_2+p_2)^h}\right)$
bi-bi reversible Hill with k-type modifier [79]	$v = V_f \frac{s_1 s_2\left(1-\frac{\Gamma}{K_{eq}}\right)(s_1+p_1)^{h-1}(s_2+p_2)^{h-1}(1+\alpha^2 m^h)}{(1+m^h)+(1+\alpha m^h)\left[(s_1+p_1)^h+(s_2+p_2)^h\right]+(1+\alpha^2 m^h)(s_1+p_1)^h(s_2+p_2)^h}$
bi-bi reversible Hill with v-type and k-type modifiers [79, 333]	$v = V_f \frac{s_1 s_2\left(1-\frac{\Gamma}{K_{eq}}\right)(s_1+p_1)^{h-1}(s_2+p_2)^{h-1}\prod_{i=1}^{n}(1+\gamma_i\alpha_i^2 m^h)}{\prod_{i=1}^{n}(1+m^h)+\prod_{i=1}^{n}(1+\alpha m^h)\left[(s_1+p_1)^h+(s_2+p_2)^h\right]+\prod_{i=1}^{n}(1+\alpha^2 m^h)(s_1+p_1)^h(s_2+p_2)^h}$
ter-bi reversible Hill [79]	$v = V_f \frac{s_{11}s_{12}s_3\left(1-\frac{\Gamma}{K_{eq}}\right)(s_{11}s_{12}+p_1)^{h-1}(s_3+p_2)^{h-1}}{\left[1+s_{11}^h+s_{12}^h+(s_{11}s_{12}+p_1)^h\right]\left[(s_3+p_2)^h+1\right]}$
bi-substrate irreversible Michaelis-Menten [84]	$v = V_f \left(\frac{S_1}{K_{m1}+S_1}\right)\left(\frac{S_2}{K_{m2}+S_2}\right)$

5 Model behaviour: key characteristics of microaerobic *E. coli* glycolysis

Summary

The steady state behaviour and control properties of models developed in the previous chapter are explored. An elementary mode analysis was performed to identify structural features of the reaction network. A supply-demand analysis around the $\frac{[ATP]}{[ADP]}$ ratio was performed to determine the distribution of flux control between glycolysis and the demand for ATP as well as the degree of free energy homoeostasis. It was shown that all three model exhibited a positive correlation between the supply flux and the $\frac{[ATP]}{[ADP]}$ ratio. Flux control by demand under aerobic conditions has been argued in publication, however the supply block controlled flux in all models. Using control analysis it was shown that the minor degree of flux control by demand on the supply block is mediated by pyruvate kinase and phosphoenolpyruvate carboxylase. A demand-stimulated increase in pyruvate kinase activity results in competition between PK/PPC and the PTS for PEP, thus reducing glucose uptake flux. The effect of short-term variations in oxygen availability on the metabolic network was explored. It was observed that the redundancy around the PEP-PYR node allowed the cell to respond to variations in O_2 in the microaerobic range by balancing acetate, ethanol and lactate production, suggesting a possible microaerobic role for LDH. Additionally the question of whether FBP can act as an internal flux sensor for transcriptional regulation is broached.

5.1 Introduction

Cellular metabolism in non-growing cells is predominantly mediated by two important moiety-conserved cycles, the energy charge which has been represented in a simplified fashion in this study as the $\frac{[ATP]}{[ADP]}$ ratio, and the redox balance or $\frac{[NADH]}{[NAD^+]}$ ratio. These two cycles are both implicated in the electron-extracting reactions of glycolysis and the TCA cycle. Under aerobic conditions, the majority of reducing equivalents (NADH) produced by glycolysis and the TCA cycle are utilised for ATP generation at the plasma membrane. The fermenting anaerobic *E. coli* cell, which relies on the substrate-level phosphorylation reactions of glycolysis for generating ATP, encounters a metabolic obstacle as NAD^+—which is required for glycolysis—needs to be regenerated in the absence of a respiratory electron acceptor. *E. coli*, like so many other micro-organisms, has solved this redox problem by reducing partially-oxidised glycolytic products to produce a range of fermentation products in various ratios to balance the redox demand of the cell (discussed in Section 1.6.3). *E. coli* exhibits unexpected metabolic behaviour in response to high substrate concentrations [54, 123, 146]: maintaining an active acetate fermentation pathway under fully-aerobic conditions, ostensibly to provide faster metabolic capabilities and thus a competitive advantage against other organisms while mitigating the lower ATP production rates of fermentation [123, 146]; and expressing a high-affinity cytochrome ($cytb_d$) under microaerobic conditions [129], potentially scavenging remaining environmental O_2 and thus offering a degree of protection to the metabolically important oxygen-sensitive enzyme PFL [132]. The particular metabolic state adopted by *E. coli* cells at any given time is thus largely constrained by redox demands, making the microaerobic state—in which it is necessary for the cell to display a degree of metabolic flexibility not required under aerobic and anaerobic conditions—an interesting mode of functioning to explore.

The usefulness of a kinetic model of central carbon metabolism in *E. coli* is displayed initially in the information that can be gleaned from *in silico* experiments. Experiments with the models developed in this study have been able to shed light on a number of key outstanding questions regarding the regulation of metabolism in *E. coli*. These questions were posed at the end of Chapter 1 and are repeated here for convenience:

1. Is the "textbook" concept of a rate-limiting step in glycolysis valid for this model organism (usually PFK)?

2. Do ATP-demanding reactions external to glycolysis control flux through glycolysis in the microaerobic state (as shown for aerobic *E. coli* [101])?

3. Is the correlation between cellular free energy and PTS activity positive [118] or negative [101]?

4. Do any particular enzymes represent key regulatory points in glycolysis, especially with regard to the cellular free energy level and redox balance?

5. Are there structural features of glycolysis in *E. coli* which provide a unique degree of flexibility in response to redox demands?

6. To what degree is *E. coli* able to respond metabolically to slight variations in levels of O_2, and how much of the O_2 adaptation requires a transcriptional response?

7. Is it possible for FBP to act as an internal glycolytic flux-sensor (as suggested by Kochanowski *et al.* [153])?

These questions are explored in the current chapter and the results of a series of *in silico* experiments are presented.

5.2 Structural features of glycolysis in fermenting *E. coli* cells

Elementary mode analysis is a useful tool for understanding important structural aspects of the topology of reaction networks [13]. A key feature of elementary mode analysis is that is a purely stoichiometric tool, and as such is independent of the reaction rate laws specified by the particular model being studied.

Elementary flux modes are non-decomposable non-overlapping "minimal set[s] of enzymes, able to operate at steady state with all irreversible reactions proceeding in the direction prescribed thermodynamically"[13]. Elementary modes identify sets of enzymes that are indispensable to the steady state system, that is, a system at steady state is inevitably generated by a combination of elementary flux modes in a particular ratio. Beyond identifying the core modes of operation present in a metabolic network, elementary mode analysis is able to identify possible futile cycling [13, 25].

To characterise the steady state behaviour of the glycolytic models developed, model O (with the succinate pathway included) was subjected to elementary mode analysis using the network analysis program Metatool [60] (as the models share the same topology, and elementary modes are structural features, it makes no difference which model is used). Import (PTS: R1-R10) and export (FRD/FRDII, ACE, ETH, LDH) reactions were set as irreversible to generate flux modes that represent a cell in mid-logarithmic growth on glucose, and to reduce the complexity of the modes identified (69 modes are identified when all reactions are reversible, 10 when the irreversibility constraint is applied). PO_4^{3-} and HCO_3^-, and the external metabolites (GLC, SUCC, ACE, ETH, LAC) were fixed. As a futile cycle involving PCK and PPC has been reported in glucose-limited chemostat cultures [122], two models were analysed, one including, and one excluding PCK. Identified elementary modes are presented in Table 5.1.

Four elementary modes (1–4) were identified in the model excluding PCK. Modes 1 and 2 represent balanced acetogenic and ethanologenic fermentation (producing 3 ATP molecules per molecule of glucose), and homolactic fermentation (producing 2 ATP molecules per molecule of glucose) respectively, demonstrating the stoichiometric constraints of the redox balance under purely

fermentative conditions. Mode 3 represents an additional fermentative route in which no flux passes through PK. Instead flux branches through the PTS (as in all glycolytic modes) and PPC. Pyruvate produced by the PTS reaction passes through PFL to acetate, generating the necessary HCO_3^- for succinate fermentation. Bypassing PK results in the loss of production of 1 ATP and only 2 ATP molecules are generated per molecule of glucose. This metabolic detour through PPC has been identified in *pykA* mutants [247]. Mode 4 represents acetogenic fermentation in the presence of O_2 in which the burden of redox balancing falls upon $cytb_d$. This mode is similar to the "overflow" metabolic state under aerobic conditions with excess glucose, in which *E. coli* redirects some of the glycolytic flux to acetate, gaining an advantage in metabolic rate while mitigating some of the loss in respiratory ATP production by generating 4 ATP per glucose molecule [123, 146].

When PCK was included in the analysis, the number of modes more than doubled. The reason for this is that the original modes were duplicated with stoichiometric readjustments to compensate for an ATP consuming futile cycle between PPC and PCK which essentially replaces the vATP reaction in the original modes. This is best illustrated with an example. The homolactic mode (2) consisted of:

$$0.5vPTS + 0.5vGLY + 0.5vPK + vLDH + \mathbf{vATP}$$

in the absence of PCK. When PCK is added both the original mode 2 and a variant exist:

$$0.5vPTS + 0.5vGLY + 0.5vPK + vLDH + \mathbf{vPPC} - \mathbf{vPCK}.$$

The difference being that ATP demand in the original has been replaced in the new mode by an ATP dissipating cycle between PPC and PCK.

Only 2 entirely new nodes were identified, modes 5 and 6 (Table 5.1). Mode 5 represents a futile cycle between $vPPC$, $vPCK$ and $vATP$ ($ATP \rightarrow ADP + PO_4^{3-}$) with no net change of intermediate concentrations, however, in isolation $vPPC$ and $vPCK$ together produce ATP which is matched by the non-specific ATP consuming reaction. When $vPPC$ and $vPCK$ cycle in the opposite direction, ATP is dissipated as described above. Mode 6 represents a high ATP-yielding (3 ATP molecules per molecule of glucose) combination of acetate and succinate fermentation in which the reversible $vPCK$ [334] is in fact active in the opposite direction to that of its typical gluconeogenic role, replacing $vPPC$ as the entry point to succinate fermentation and in so doing, generating an extra ATP molecule. The potential of mode 6 as a route for engineering increased succinate producers with concomitant ATP generation in *E. coli* has been realised in *ppc* mutant strains with increased levels of PCK [335].

Using elementary mode analysis, it has been shown how *E. coli* is able to respond with versatility to changing redox and energy requirements. Additionally, futile cycling introduced by PCK provides possibly even greater flexibility with regard to managing the cellular energy charge and producing biosynthetic equivalents via the succinate fermentation pathway.

Table 5.1: Elementary modes of glycolysis under microaerobic fermentative conditions. Elementary flux modes were calculated using Metatool [60]. The PTS reaction, fermentative end steps, including FRD, ACE, ETH, LDH, $vNADH$, and the non-specific ATP consuming reactions were set as irreversible. Initially PCK was excluded, which generated modes 1 to 4. Upon inclusion of PCK, with the exception of stoichiometric variations of the original four modes, the futile cycling mode 5 and the reverse PCK mode 6 appeared. For convenience, reactions have been grouped into pathways.

	net reaction	constituent steps
1:	$GLC = 2CO_3 + ACE + ETH$	vPTS[a] + vGLY[b] + vPK + 2vPFL + vETH + vACE + 3vATP
2:	$GLC = 2LAC$	vPTS + vGLY + vPK + 2vLDH + 2vATP
3:	$GLC = SUCC + ACE$	vPTS + vGLY + vPPC + vSUCC[c] + vPFL + vACE + 2vATP
4:	$GLC = 2CO_3 + 2ACE$	vPTS + vGLY + vPK + 2vPFL + 2vACE + 2vNADH + 4vATP
5:	futile cycle	$vATP - vPPC + vPCK$[d]
6:	$GLC = SUCC + ACE$	vPTS + vGLY + vPCK + vSUCC + vPFL + vACE + 3vATP

[a] $vPTS = R1 + R2 + R3 + R4 + R5 + R6 + R7 + R8 + R9 + R10$

[b] $vGLY = vPGI + vPFK + vTPI + 2vGAPDH + 2vPGK + 2vPGM + 2vENO$

[c] $vSUCC = vMDH + vFUM + vFRD + vFRDII$

[d] For the purpose of this analysis, HCO_3^- was included in the stoichiometry of the PCK reaction. It is excluded and fixed in the original model.

5.3 Stability analysis

An updated diagram of the full glycolytic model with succinate fermentation is shown in Fig. 5.1. In the figure, reactions have been rearranged around the $\frac{[ATP]}{[ADP]}$ ratio to better visualise energy dynamics in the network.

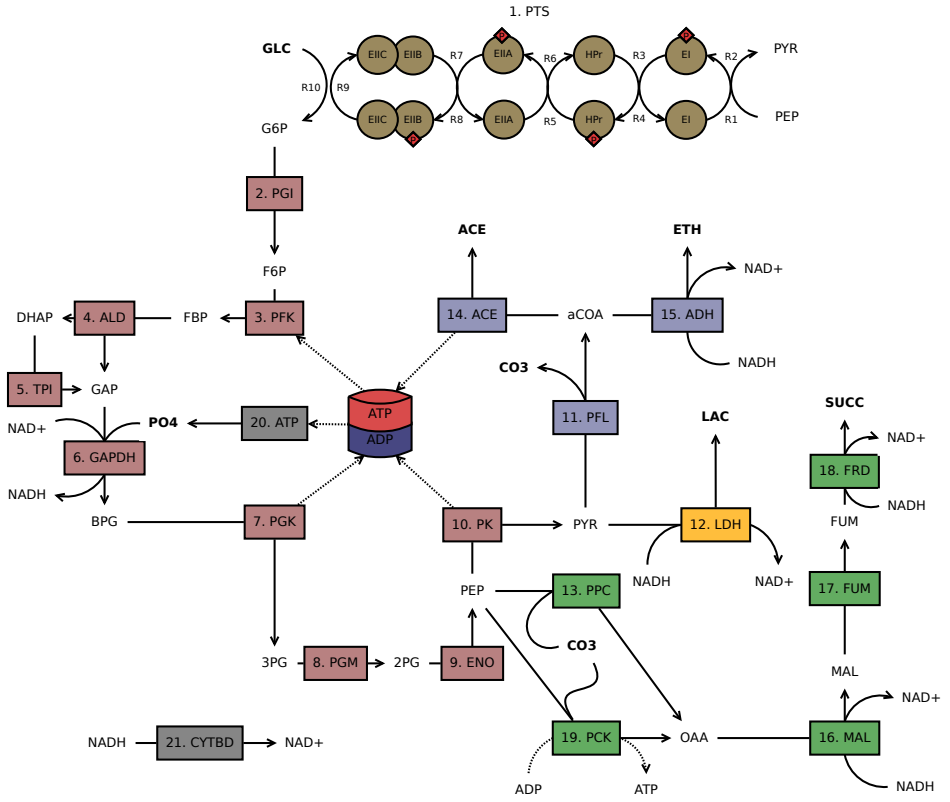


Figure 5.1: A circular representation of the topology of the succinate containing models arranged with $\frac{[ATP]}{[ADP]}$ at the centre. Reactions lines pointing towards the $\frac{[ATP]}{[ADP]}$ ratio in the middle of the diagram contribute to the $\frac{[ATP]}{[ADP]}$ ratio (i.e. by generating ATP); those pointing away consume ATP. PCK is an exception to this arrangement and is placed at the bottom of the image as it exhibits significantly lower steady state flux in all models, thus only weakly affecting the $\frac{[ATP]}{[ADP]}$ ratio. Central carbon pathway arrows do not indicate thermodynamically favoured directions but rather the direction of flux during fermentative steady state (the exception is PCK in model Q which proceeds in reverse). PTS – brown blocks, Glycolysis – mauve blocks; succinate fermentation – green blocks; lactate fermentation/LDH – amber block; acetate/ethanol fermentation – light blue blocks; ATP consumption and cytb_d – grey blocks. Reactions involving $\frac{[ATP]}{[ADP]}$ are indicated by a dotted line. Fixed metabolites are in bold.

Before further characterising the regulatory behaviour of the constructed models in response to perturbations by performing metabolic control analysis, the stability of the three models was analysed. The stability of a dynamical system relates to the asymptotic behaviour of the system over time in response to small perturbations in the state of the system. Equilibrium points (or

critical points) refer to positions in the phase space of a system which do not change over time, that is, they are constant solutions of a system [336]. A method of elucidating the dynamical behaviour around a critical point of a complex non-linear system of ODEs is to instead analyse a linear approximation of the neighbourhood of the system in the proximity of the critical point. This can be usefully represented by the Jacobian matrix (a matrix of first-order partial derivatives [336]) at the critical point. The eigenvalues of the Jacobian matrix at a critical point characterise the behaviour and stability of the system around this point [337]: if the real parts of the eigenvalues of the Jacobian matrix are negative, the critical point is stable and small perturbations in the state will return the system to the original state; if however there is at least one eigenvalue with a positive real part, the critical point is unstable, and thus, there exists a set of perturbations to the state of the system at the critical point which will cause the system to destabilise [336]. The eigenvalues of the Jacobian matrix for the steady states of the constructed models were calculated and were all shown to be purely real and negative. Thus the steady states (i.e. critical points) presented in the models are stable.

5.4 Metabolic control analysis

Metabolic control analysis was performed on the three full models (including succinate) to derive the steady state flux and concentration control coefficients. Flux and concentration control coefficients for the three models are plotted in Fig. 5.2. The top four flux and $\frac{[ATP]}{[ADP]}$ concentration control coefficients for the three models are presented in Table 5.2 (note that only the full succinate including models are tabulated as the control properties of the simpler succinate excluding models were almost identical). When a reaction reference is italicised or preceded by a v, the reaction itself is being referred to in the context of the reaction network, and not the enzyme responsible for catalysis.

In all models, the PTS components dominated the control of glycolytic flux (also see Fig. 5.2). $R8$ contributed the highest degree of control, and second to $R8$ a significant degree of flux control was exerted by $R10$. This result is not unexpected as the PTS enzyme with the greatest effect on PTS flux has been shown experimentally to be the transporter EIICB ($R_{IICB}^{J_{PTS}} = 0.7$ [338]) by varying protein concentration. This was replicated in the isolated PTS model ($R_{IICB}^{J_{PTS}} = 0.9$ [68]). In the isolated PTS model, reaction $R8$ exercises the most flux control with $C_{R8}^{J_{PTS}} = 0.6$. $R8$ is the reaction in which EIIA dissociates from phosphorylated EIICB. $R10$ is the glucose transport step via EIICB. Both of these reactions are expected to exhibit a high degree of flux control by virtue of their association with the membrane bound EIICB. Additionally, the reaction with the third highest degree of flux control ($R7$) was also associated with EIICB. Interestingly however, $R3$, the association/dissociation reaction between phosphorylated EI and HPr, also demonstrates a degree of flux control in models O and P. Steady state PYR concentrations are similar between the models but the steady state PEP concentration achieved by model Q is significantly lower. It

Table 5.2: The top four flux and $\frac{[ATP]}{[ADP]}$ concentration control coefficients of models O, P and Q. Flux control coefficients are calculated according to the flux through $vPGI$ representing glycolysis.

flux		$\frac{[ATP]}{[ADP]}$	
model O			
$C_{R8}^{J_{PGI}}$	0.56	C_{vATP}^{RA}	−2.50
$C_{R10}^{J_{PGI}}$	0.27	C_{R8}^{RA}	1.56
$C_{R7}^{J_{PGI}}$	0.09	C_{R10}^{RA}	0.75
$C_{R3}^{J_{PGI}}$	0.03	C_{vETH}^{RA}	−0.29
model P			
$C_{R8}^{J_{PGI}}$	0.56	C_{vATP}^{RA}	−2.54
$C_{R10}^{J_{PGI}}$	0.27	C_{R8}^{RA}	1.56
$C_{R7}^{J_{PGI}}$	0.10	C_{R10}^{RA}	0.74
$C_{R3}^{J_{PGI}}$	0.03	C_{vETH}^{RA}	−0.30
model Q			
$C_{R8}^{J_{PGI}}$	0.59	C_{vATP}^{RA}	−2.46
$C_{R10}^{J_{PGI}}$	0.26	C_{R8}^{RA}	1.38
$C_{R7}^{J_{PGI}}$	0.14	C_{R10}^{RA}	0.62
$C_{vPK}^{J_{PGI}}$	−0.09	C_{R7}^{RA}	0.32

appears that the low PEP concentration is a result of the maximal forward rate of vPK in model Q being much higher than in the other models. By virtue of the greater degree of competition exhibited by vPK for PEP, vPK has a small degree of flux control. The general control patterns of the models immediately suggest that glycolytic flux is almost entirely controlled by reactions that supply glycolytic intermediates, and as such, the supply reactions are unlikely to be able to respond to increased demand for ATP (as in the aerobic situation observed by Koebmann *et al.* [101]).

In all three models, the flux most responsive to changes in the rate of controlling reactions was that of $vLDH$ (O: $C_{R8}^{J_{LDH}} = 3.74$, P: $C_{R8}^{J_{LDH}} = 4.22$, Q: $C_{R8}^{J_{LDH}} = 3.60$). This is likely due primarily to the high degree of cooperative binding LDH displays towards PYR ($h_{LDH} = 4$), and may indicate that LDH plays an important role in glycolysis by virtue of its ability to respond rapidly to changes in PYR concentration and NADH, and thus glycolytic flux in general.

As could be expected, $\frac{[ATP]}{[ADP]}$ control displays similar control properties to the glycolytic flux as it is positively regulated by $R8$ and displays a negative control coefficient towards the ATP demand reaction. Once more, only one non-PTS reaction beside $vATP$ appears in the top four control coefficients. $vETH$ exhibits a moderate degree of control over the $\frac{[ATP]}{[ADP]}$ ratio in models O and P. As the control coefficients are negative, it appears that $vETH$ exerts its control over the $\frac{[ATP]}{[ADP]}$ ratio by competing with $vACE$, an ATP-generating reaction which is not intrinsic to the

core glycolytic pathway. Being thus uncoupled from glycolytic flux, to a degree, all acetogenic ATP is a "bonus" contributing to the $\frac{[ATP]}{[ADP]}$ ratio. These control features will be discussed further in the following sections pertaining to the behaviour of the models in terms of $\frac{[ATP]}{[ADP]}$ and $\frac{[NADH]}{[NAD^+]}$ in response to changes in supply and demand for energy, and changes in O_2 availability.

5.5 Dealing with variations in ATP demand

A useful paradigm for considering the energy economy of the cell is supply and demand analysis (SDA) [98] (Section 1.5) which emerged out of the MCA framework. SDA involves separating a network of reactions into two blocks, a block that supplies a particular intermediate, and a block that demands or consumes the intermediate. In the ideal case, the identified intermediate is the only means of communication between the two blocks. A consequence of the summation and connectivity theorems of MCA (Section 1.5) is that control coefficients can be expressed purely in terms of elasticities, and thus supply and demand block elasticities can be used to calculate the flux- and concentration control coefficients of each block according to equations 5.1-5.4 [15] reproduced here from Chapter 1 for convenience:

$$C^J_{supply} = \frac{\varepsilon^{demand}_s}{\varepsilon^{demand}_s - \varepsilon^{supply}_s} \quad (5.1) \qquad C^s_{supply} = \frac{1}{\varepsilon^{demand}_s - \varepsilon^{supply}_s} \quad (5.3)$$

$$C^J_{demand} = \frac{-\varepsilon^{supply}_s}{\varepsilon^{demand}_s - \varepsilon^{supply}_s} \quad (5.2) \qquad C^s_{demand} = \frac{-1}{\varepsilon^{demand}_s - \varepsilon^{supply}_s} \quad (5.4)$$

where J is a steady state flux and s is a steady state metabolite linking the supply and demand blocks.

A key realisation of this process is that flux control is mutually exclusive, resulting in the potential of functional differentiation between the supply and demand blocks; as one block assumes control of the flux through the system, the other relinquishes flux control and becomes the determining block of the degree of homeostatic control over the concentration of the intervening metabolite [98]. SDA has been generalised to analyse all the metabolites in a network, identifying regulatory metabolites and quantifying the degree of functional differentiation around these metabolites [339].

Glycolytic flux in *E. coli* has been argued to be controlled by the demand for ATP [101] under aerobic respiratory conditions. However, the regulation of the PTS is not fully understood, and many contradictory responses have been observed to changes in the $\frac{[ATP]}{[ADP]}$ ratio, proton motive force, respiratory activity and the $\frac{[PEP]}{[PYR]}$ ratio (Section 4.1). The salient study already mentioned [101] showed that ATP demand controls flux in *E. coli*, that is, increases in the demand for ATP reduced the $\frac{[ATP]}{[ADP]}$ ratio, which in turn initiated an increase in the ATP-supplying flux through glycolysis. Thus, a negative correlation exists between the $\frac{[ATP]}{[ADP]}$ ratio and demand-altered glycolytic flux. In that study, demand was calculated to exert at least 75% of control over glycolytic flux ($C^J_{dem} \geq 0.75$), marginalising the control of the supplying reactions. An alternate study performed on succinate-

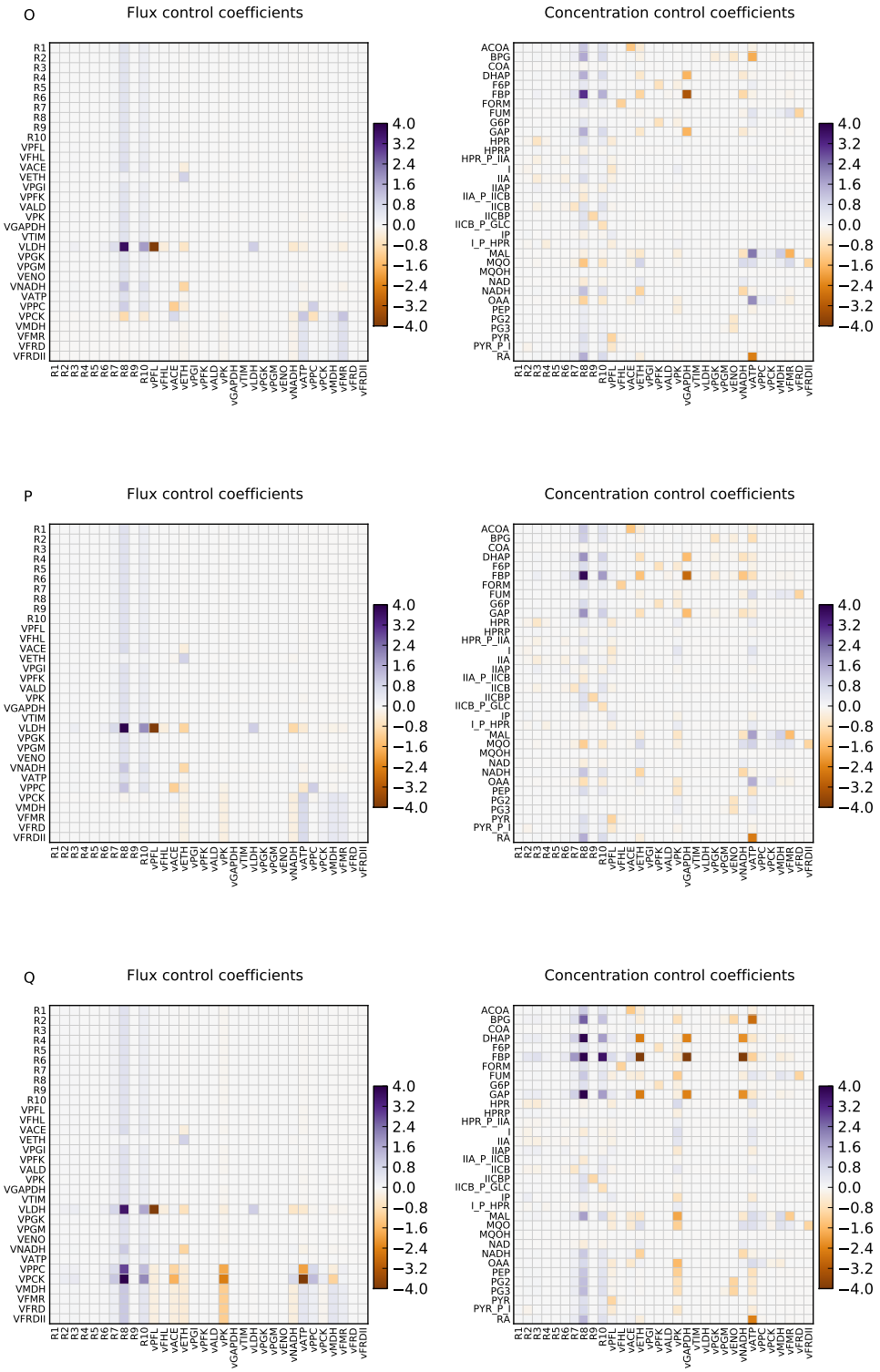


Figure 5.2: Flux (left) and concentration (right) control coefficients for the reactions of models O, P and Q including the succinate pathway. Controlling reactions are indicated on the x-axes, dependent fluxes and metabolite concentrations are indicated on the y-axes. Colorbars are normalised to represent identical scales, and for visual clarity are saturated over 4 and below -4.

grown cells [118] observed that upon the addition of uncouplers, respiratory inhibitors, or the modulation of ATP synthase activity, invariably reduction of the $\frac{[ATP]}{[ADP]}$ ratio in the cell resulted in a reduction in α-MG uptake. Thus, a positive correlation between $\frac{[ATP]}{[ADP]}$ ratio and PTS activity was observed. Also a study using weak acid perturbation of the ATP demand block in aerobic cells generally resulted in a reduced growth rate (i.e. ATP demand) and no change in glucose uptake rate [215], suggesting that the supply block exercised the majority control over glycolytic flux. Though EIICB is known to exert the majority of control over PTS activity [338], it has been shown to exert almost no control over glycolytic flux (measured as glucose oxidation) under aerobic conditions [268], though modulating the concentration of EIICB even slightly below the wild type level shifts the control of flux towards EIICB. This demonstrates a possible flexibility in the role of EIICB in regulation of glycolytic flux.

To characterise the regulatory response of the constructed models to variations in the supply of and demand for ATP, the models were divided into two reaction blocks. As glycolysis produces a net increase in the $\frac{[ATP]}{[ADP]}$ ratio, for each model, the whole set of $\frac{[ATP]}{[ADP]}$-dependent reactions of glycolysis was treated as the supply block (i.e. $vPFK$, $vPGK$, vPK, $vPCK$ and $vACE$, reactions 3, 7, 10, 19 and 14 in Fig. 5.1); $vATP$ alone was designated as the demand block. In this division the only means of communication between the supply and demand blocks is the $\frac{[ATP]}{[ADP]}$ ratio.

SDA was performed by fixing and scanning the $\frac{[ATP]}{[ADP]}$ ratio (RA) over a range around the steady state from 10–1000 %, computing a new steady state flux at each point to generate a "rate characteristic". The rate characteristics for models O, P and Q are shown in Fig. 5.3. Using the control properties of a model, a "block elasticity" [98]—conceptually similar to the response coefficient of the flux of a block to changes in the intervening metabolite—can be calculated and the distribution of flux control determined. The partitioned response property of control analysis [11] ($R^J_x = \varepsilon^i_x C^J_i$, where x is a parameter or variable metabolite, i is an individual reaction, J is the state flux, and R, ε and C are the response coefficient, elasticity and control coefficient respectively), allows the block response coefficient (or elasticity) to be calculated from constituent partial response coefficients [15, 339]. In this context the block elasticity for supply is:

$$\varepsilon^{J_s}_{RA} = {}^{vPFK}R^{J_s}_{RA} + {}^{vPGK}R^{J_s}_{RA} + {}^{vPK}R^{J_s}_{RA} + {}^{vPCK}R^{J_s}_{RA} + {}^{vACE}R^{J_s}_{RA} \tag{5.5}$$

$$= \varepsilon^{vPFK}_{RA} C^{J_s}_{vPFK} + \varepsilon^{vPGK}_{RA} C^{J_s}_{vPGK} + \varepsilon^{vPK}_{RA} C^{J_s}_{vPK} + \varepsilon^{vPCK}_{RA} C^{J_s}_{vPCK} + \varepsilon^{vACE}_{RA} C^{J_s}_{vACE} \tag{5.6}$$

and the block elasticity for demand is:

$$\varepsilon^{J_d}_{RA} = \varepsilon^{vATP}_{RA} C^{J_d}_{vATP} \tag{5.7}$$

where J_s and J_d are the steady state fluxes of the supply block and demand block respectively.

Each constituent partial response coefficient is the product of the elasticity of that particular reaction to RA and the control coefficient of that reaction on the block flux. The partial response coefficients quantify the contribution of each reaction to the total response and thus can be used to identify the reactions responsible for mediating the dynamic behaviour of each block. Note that the demand block only consists of one reaction, and thus one response coefficient. Block elasticities and partial response coefficients for models O, P and Q, as well as calculated flux and concentration control coefficients are shown in Table 5.3:

Table 5.3: Supply and demand analysis of models O, P and Q. Block elasticities (bold), partial response coefficients and control coefficients are represented.

model	$\varepsilon^{J_s}_{RA}$	$^{vPFK}R^{J_s}_{RA}$	$^{vPGK}R^{J_s}_{RA}$	$^{vPK}R^{J_s}_{RA}$	$^{vPCK}R^{J_s}_{RA}$	$^{vACE}R^{J_s}_{RA}$	$\varepsilon^{J_d}_{RA}$
O	**0.0059**	4.7×10^{-5}	0	0.0007	0.006	-0.001	**0.407**
P	**0.013**	2.6×10^{-5}	3.2×10^{-7}	0.0047	0.0089	-0.0005	**0.407**
Q	**-0.0003**	4.7×10^{-6}	0	0.0030	-0.0007	-0.003	**0.407**
	C^J_{supply}	C^J_{demand}	C^{RA}_{supply}	C^{RA}_{demand}			
O	1.02	-0.02	2.49	-2.49			
P	1.03	-0.03	2.54	-2.54			
Q	1.0	0.0	2.45	-2.45			

In contrast to the results in the aerobic glucose studies mentioned above [101, 268], but in agreement with the study by Salmond *et al.* [215] ATP supply, and not demand, controlled flux in all models; the demand block elasticity exceeding the absolute value of the supply block elasticity by an order of magnitude. Indeed, the supply block elasticities were more similar to the demand block elasticity (0.04) calculated by Koebmann *et al.* [101]. The demand block elasticity is indicative of a simplified representation of ATP demand ($vATP$ is a simple mass-action equation) and thus changing the rate law to an equation with significant saturation effects would reduce the demand block elasticity significantly. However, there is a significant degree of saturability inherent in the fact that ATP is part of a moiety-conserved cycle subject to mass conservation ($ATP = S\frac{RA}{RA+1}$—where $S = ATP + ADP$ and $RA = \frac{ATP}{ADP}$—conforms to the same structure as the saturable Michaelis-Menten rate equation). This naturally does not resolve the degree of saturation present in the demand block which would be made explicit by a saturable rate equation. Assuming a demand block elasticity of 0.04, as determined by Koebmann *et al.* [101], would imply that the supply and demand blocks share control almost equally. The wildtype EIICB concentration is located nearby a transition point between demand and supply control of flux [268]). It is thus also conceivable that reduction of the EIICB concentration under aerobic conditions may shift flux control to the demand block.

Additionally, the supply block elasticities for models O and P were slightly positive. This disagrees with the results produced by Koebmann *et al.* [101] but agrees with the observations of the succinate grown cells by Rohwer *et al.* [118]. It may be that the difference in control properties

between model Q and the observations of Koebmann *et al.* [101], is due to the fermentative growth conditions represented by the models (vs. the aerobic experimental data in [101]). If the PTS can be up-regulated via the glycolytic network in a way that responds to increases in demand flux and the concomitant decrease in the $\frac{[ATP]}{[ADP]}$ ratio, it is not by a simple route. The only internal metabolic regulatory routes that affect the PTS are via PEP and PYR. PEP is intrinsically linked to the PTS as both substrate and eventual product of the linear glycolytic pathway, thus increasing PTS activity is likely to increase PEP concentration and *vice versa*, potentially producing a feed-forward loop, which explains the positive supply block elasticity. Altering the kinetics of lower glycolysis, especially PK and the enzymes leading to PEP may increase PEP concentrations and thus activate the PTS. Reducing PYR concentrations may be achieved by an increase in acetate fermentation in response to increased ADP. Ultimately it may be that the observed membrane effects on the PTS during respiration are modulating the pathway activity (Section 4.1). The absence of a significant proton motive force across the cell membrane, which has been argued to be a significant determinant of PTS activity may be playing a role.

An implication of equations 5.3 and 5.4 is that the degree of homeostatic control that a supply and demand system exercises over the concentration of the intervening metabolite is inversely proportional to the sum of the block elasticities; when one block elasticity tends towards 0, and thus assumes control over flux, the other elasticity determines the concentration range over which the metabolite is maintained. Thus, the steeper the block elasticities, the less control is exercised by the blocks over metabolite concentration, and the more constrained the metabolite is. In the developed models the sum of the block elasticities was lower than in [101], resulting in fourfold greater concentration control coefficients for the two blocks, indicating a broader region over which the $\frac{[ATP]}{[ADP]}$ ratio is maintained. Assuming the demand block elasticity determined by [101] of 0.04 would result in significantly higher concentration control coefficients, and thus an even broader range for the $\frac{[ATP]}{[ADP]}$ ratio. The constituent partial response coefficients of the supply block characterise the points of flux control in response to $\frac{[ATP]}{[ADP]}$ variation. Though relatively low due to the dominance of the supply block control over flux, the greatest partial response coefficient in the supply blocks of models O and P was that of PK (a significantly lower supply block partial response coefficient for vPK was observed in model Q and the supply curve gradient is very shallow). The flux response to changes in supply or demand flux and the $\frac{[ATP]}{[ADP]}$ ratio is explored further.

SDA rate characteristics provide a useful means of visualising the response of a system to changes in demand or supply rate. In the case of the current models (Fig. 5.3), an increase in the rate of the demand block (easily visualised as a shifting of the demand curve upward) produces a new steady state with a lower $\frac{[ATP]}{[ADP]}$ ratio due to the unresponsiveness of the supply block, and small changes in flux. What is clear from the outset is that at saturating (and fixed) external glucose concentrations, the only means of altering PTS flux (a necessary precondition to achieving a different glycolytic flux) is by altering the concentration of either PEP or PYR (the K_{eq} towards G6P formation is 5.9×10^5). Thus, an explanation of the supply block response to increased demand

under saturating glucose concentrations has to involve a chain of causality that includes a decrease in steady state PEP or PYR concentrations. The only significant partial response coefficients in the supply blocks of models O and P were those of PCK and PK, contributing almost the total block response, and thus identifying these two enzymes as the regulatory entry point for the demand block to exert an effect on the supply flux (albeit a minor effect in a system where supply exerts the majority of flux control). The positive PCK and PK partial response coefficients are the product of negative elasticities towards $\frac{[ATP]}{[ADP]}$—an increase in the $\frac{[ATP]}{[ADP]}$ ratio is a decrease in the substrate (ADP) and an increase in product (ATP) of these reactions which will inhibit them—and a negative supply block flux control coefficient, inevitably due to the competition between PCK and PK, and the PTS for PEP. The implication of this is that the increase in demand for ATP immediately lowers the $\frac{[ATP]}{[ADP]}$ ratio, a change which is transmitted to the supply block via activation of PCK and PK, and competition for PEP slightly inhibits the PTS and thus glycolytic flux, producing less ATP per mole of glucose resulting in a lower steady state $\frac{[ATP]}{[ADP]}$ ratio. PCK and PK would be expected to play a more significant role if indeed the demand block elasticity were lower and flux control shared between the two blocks.

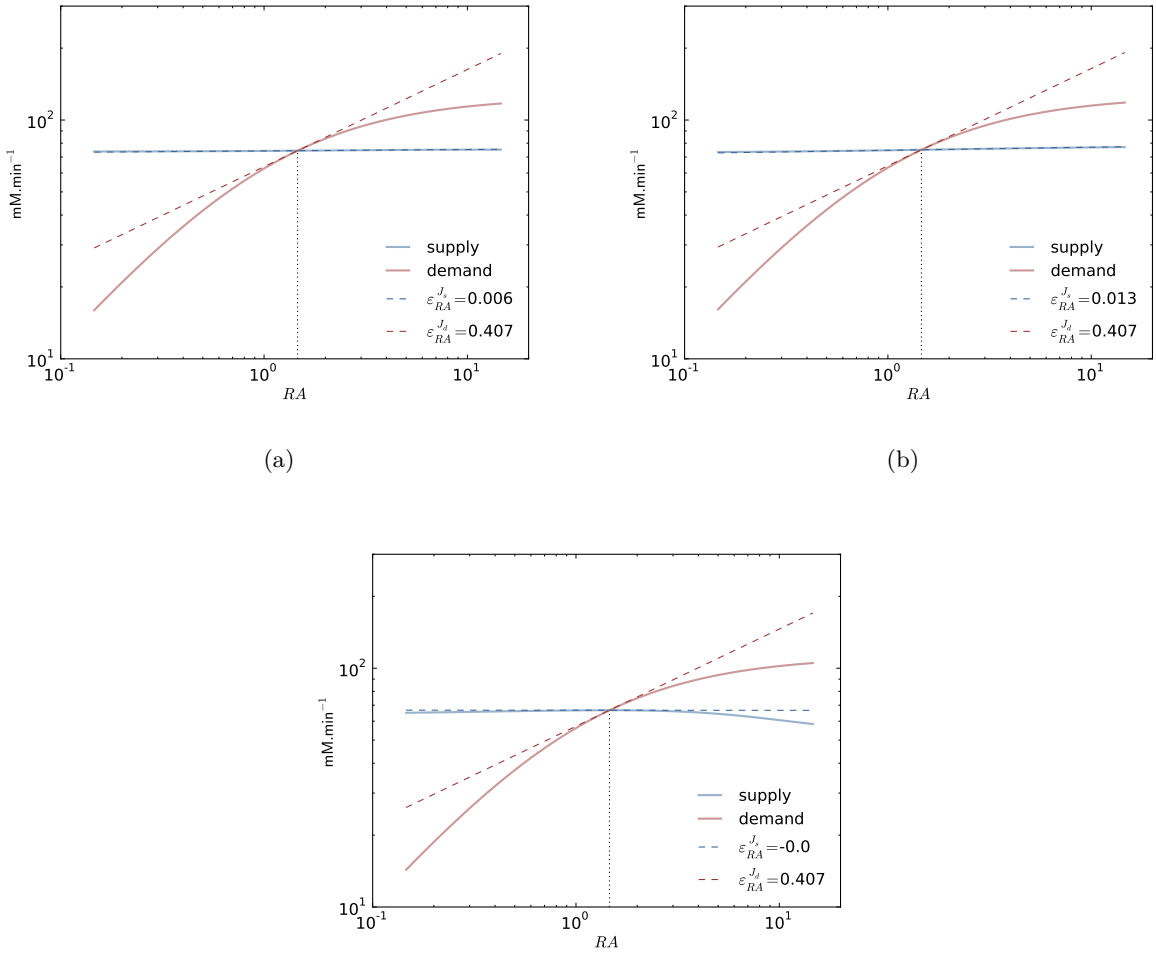


Figure 5.3: Supply and demand rate characteristics around the $\frac{[ATP]}{[ADP]}$ ratio (RA) in models (a) O, (b) P and (c) Q. The supply blocks consist of $vPFK$, $vPGK$, vPK, $vPCK$ and $vACE$; the demand block is a single reaction $vATP$, representing the ATP demand. The $\frac{[ATP]}{[ADP]}$ ratio was varied over a range of 10–1000 % around the steady state value of 1.46. The $\frac{[ATP]}{[ADP]}$ ratios and the resultant steady state fluxes are plotted on the x and y axes respectively (axes are logarithmic). The steady state point is marked with a dotted line. ε values (representing the gradient of each curve) are "block elasticities" [98] that are conceptually similar to response coefficients, quantifying the response in block flux to changes in the concentration of the intermediate metabolite, in this instance $\frac{[ATP]}{[ADP]}$.

Fig. 5.4 explores this relationship, showing a the relative changes in steady state flux of models O, P and Q over a ±20 % change in ATP demand (by modulating k_{ATP}). It is immediately clear that the glucose uptake flux (J_{R9}) and consequently the ethanol and acetate fluxes exhibited a similar but minor relative change in the three models in response to changing the ATP demand, while the largest relative response was seen in the LDH flux. This flexibility may be mediated

by the high substrate binding cooperativity of LDH and bears implications for redox balancing, which are discussed below. The increase in LDH activity with supply flux also suggests that upon a sudden increase in glycolytic flux, the relative proportion of lactate produced should increase.

5.6 FBP as flux sensor

The increase in FBP concentration with glycolytic flux (Fig. 5.4) has been suggested as a mechanism of internal flux-sensing by *E. coli* [153, 340], enabling a flux-dependent transcriptional response via its inhibition of the *Cra* transcription factor, a global regulator of glycolytic and gluconeogenic genes [341]. *Cra* generally promotes gluconeogenesis and inhibits key glycolytic genes [341]. This concept describes a generic flux sensor that acts at the metabolic level, thus enabling the cell to respond transcriptionally to a wide range of extracellular signals (e.g. changes in external metabolites) in a fashion that does not require extracellular sensory machinery. The mechanism agrees in principle with the control properties of the models developed in this study. Hypothetically, an increase in glucose uptake and thus the flux-controlling PTS activity, results in an increased FBP concentration and glycolytic flux. FBP exercises its inhibitory effect on *Cra*, resulting in an upregulation of the expression of glycolytic genes to reconfigure cellular metabolism towards higher flux. Indeed the most significant positive control coefficients promoting FBP concentration are to be found within the PTS (Fig. 5.2), thus describing the degree to which FBP and glycolytic flux are linked. In a core modelling approach it was shown that, by virtue of the feed-forward allosteric effect of FBP on PK, it is possible for FBP concentrations to increase linearly over a range far exceeding the K_m of the FBP consuming ALD reaction [153]. A linear increase in FBP concentration with flux, and linear decrease in *Cra* activity with flux were both observed experimentally [153]. The linear relationship between FBP and flux was sustained up to four times the ALD K_m in aerobic chemostat cultures, and was suggested as possibly being causal.

Fig. 5.5 shows the steady state response of the FBP concentration toward alteration of the supply flux in models O, P and Q. As EIICB concentration exhibits the highest flux response coefficient of the PTS components [68], this component was modulated to simulate an altered PTS activity. FBP concentration increased non-linearly in response to flux in all three models, increasing significantly more in model Q. The lowest increase in FBP was exhibited by model P. Though none of the FBP curves were linear near the steady state flux of 22 mM.min^{-1}, the curve that displayed the smallest change in gradient was that of model P. This was not an unexpected result in the context of the study by Kochanowski *et al.* [153], as the steady state FBP concentration was far beyond saturating in models O and Q, but not in P. Thus only model P exhibited the ability to significantly modify PK activity in response to FBP fluctuations. Though FBP promisingly increases with flux in these models, whether the models in fact support the idea of FBP acting as a flux sensor in *E. coli* in the manner suggested by Kochanowski *et al.* [153], is dependent upon the degree of linearity required between glycolytic flux and FBP concentration.

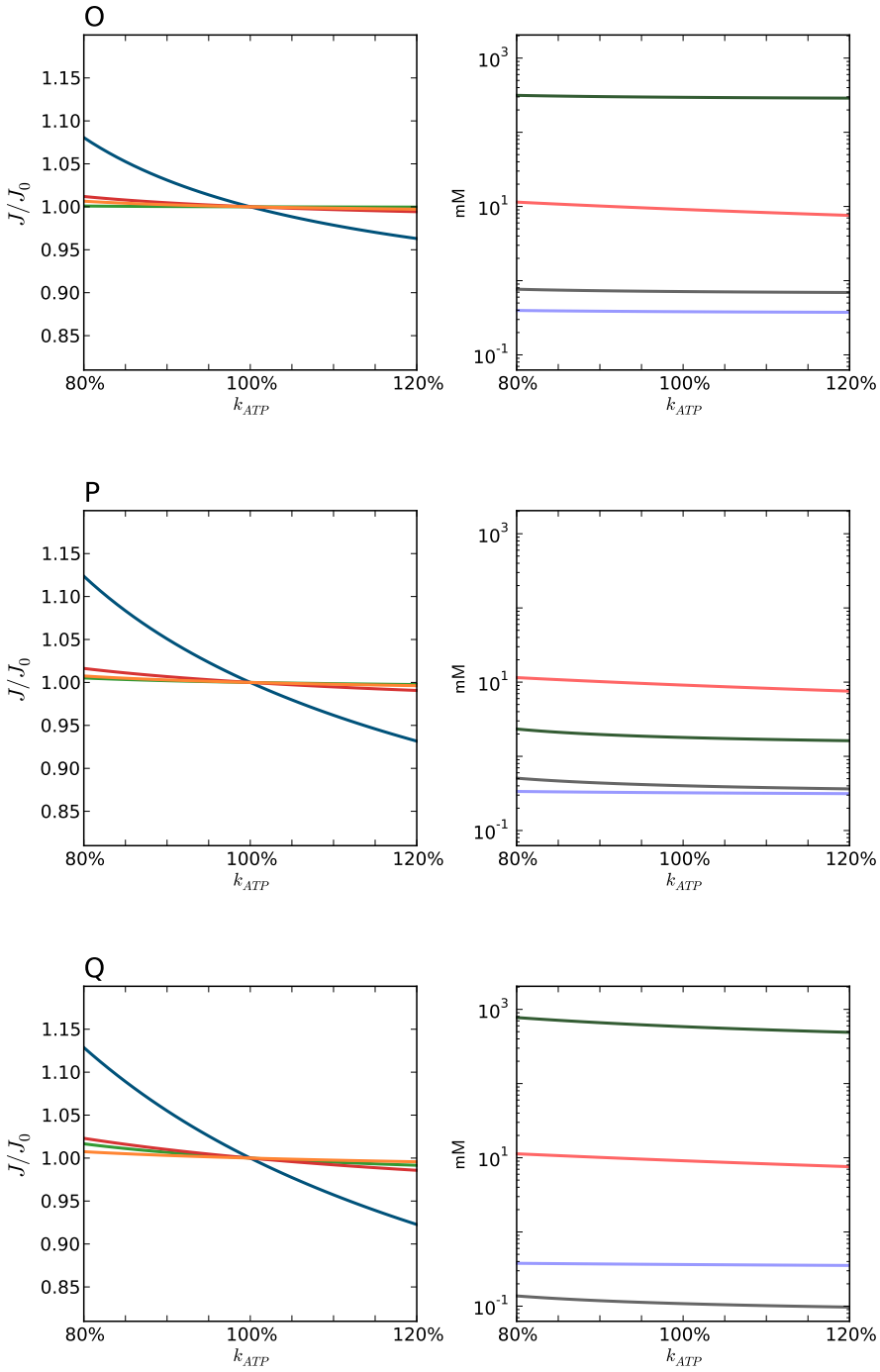


Figure 5.4: Steady state response to perturbations in ATP demand. Steady state variables were evaluated over an ATP demand range of ±20 % in models O, P and Q. The steady state flux (J) is plotted in the left hand block, and is normalised to indicate the relative change flux with respect to the steady state value, i.e. $\frac{J}{J_0}$, where J_0 is the respective reference steady state flux for each reaction (J_{R9} ——, J_{vLDH} ——, J_{vACE} ——, J_{vETH} ——). Concentrations are plotted on the right ($\frac{[NADH]}{[NAD^+]}$ ——, $\frac{[ATP]}{[ADP]}$ ——, FBP ——, PEP ——). The right y-axis is log-scaled.

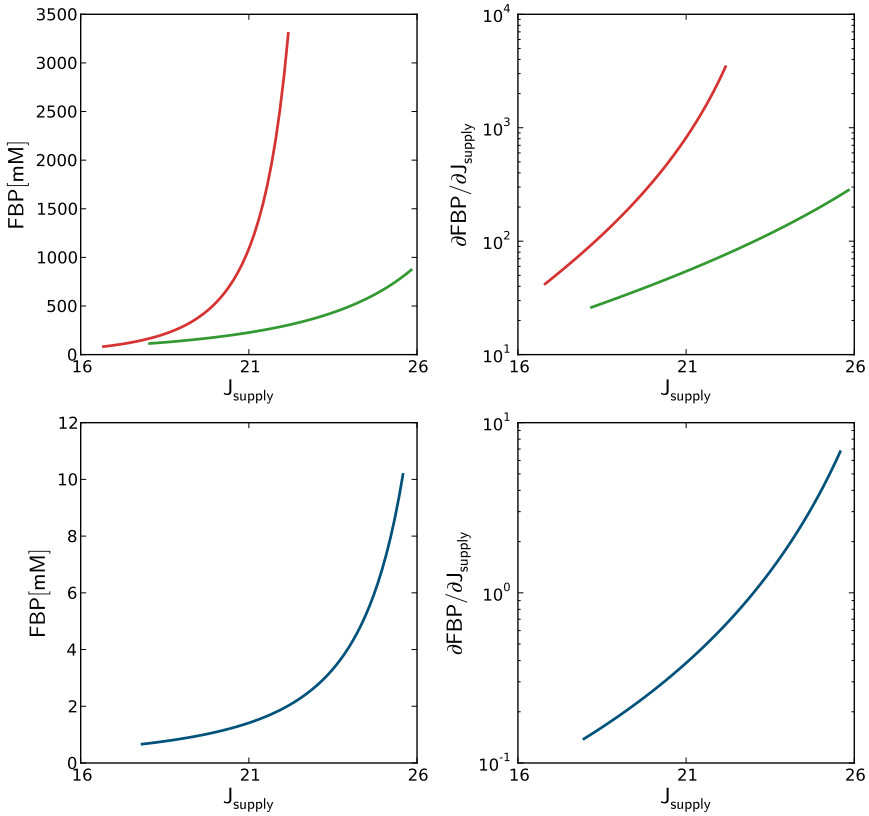


Figure 5.5: Steady state response in FBP concentration to perturbation of the supply flux. The EIICB moiety concentration was perturbed in models O (—), P (—) and Q (—) to produce an altered PTS activity. Left blocks — steady state FBP concentration is plotted against glycolytic supply flux (J_{supply}, in mM.min^{-1}, as represented by $R8$), right blocks — the relative change in FBP for a change in flux (i.e. the gradient) is plotted on a logarithmic scale.

5.7 Dealing with variations in O_2 availability

A number of salient facts have emerged from analysis of the flux responses of *E. coli* mutants to variations in O_2. A competition exists *in vivo* at the pyruvate branch point between the PFL and LDH reactions, where PFL dominates under low O_2 conditions producing acetate, ethanol and formate, in greater abundance than lactate [129, 132, 260–263] (see Fig. 4.11). Under typical microaerobic conditions the acetate flux will dominate ethanol flux due to the lower activity of ADH [150] and the activity of cytb$_d$ [132] which shares redox balancing responsibilities with ADH. This evolutionary behaviour, which is likewise displayed by the models produced in this study,

generates extra ATP for the cell by coupling ADP phosphorylation to acetate fermentation and provides greater flexibility when stoichiometrically balancing redox requirements. As the cells move to purely anaerobic conditions, acetate and ethanol flux equalise and LDH flux is marginalised, representing a stoichiometric reoxidation of the NADH produced by glycolysis in the absence of $cytb_d$ activity [132]. Glycolysis during growth on glucose produces two reducing equivalents which are stoichiometrically reoxidised by either homolactic fermentation, or an equal flux through acetate and ethanol [129, 132], or a combination of these two pathways [261–263]. There is built-in redundancy in the microaerobic fermentation routes of *E. coli* as *pflA* mutants have been shown to grow on glucose by homolactic fermentation [279]. It is also worth bearing in mind that the high-affinity $cytb_d$ is expressed under microaerobic conditions and has been argued to protect the O_2-sensitive PFL by "mopping up" excess O_2 [132].

To simulate changing O_2 availability, the maximal rate of the $cytb_d$ reaction (V_{fNADH}) was varied in model P at steady state by ±20 %, assuming that the $cytb_d$ reaction is not saturated. The behaviour or all models was similar and model P alone is illustrated in Fig. 5.6a. The ubiquitous response is that changes in O_2 availability are compensated for by balancing acetate and ethanol flux. In response to decreasing O_2, the $\frac{[NADH]}{[NAD^+]}$ ratio increases as $cytb_d$ is unable to oxidise NADH, and acetate flux decreases while the oxidising ethanol flux increases in response to the increased redox burden. This behaviour occurs even though ADH has a much lower maximal activity than the acetate enzymes [150, 260], and it has been observed in cells that are grown under decreasing O_2 availabilities [132]; as growing cells adapt to increasing anaerobiosis the $\frac{[NADH]}{[NAD^+]}$ ratio increases from about 0.1-0.4 under microaerobic conditions, and up to 0.8 and higher under fully anaerobic conditions [132, 279]. Though these observations were made in growing cultures able to respond transcriptionally, this trend can be seen in the model simulation of a non-growing culture as the $\frac{[NADH]}{[NAD^+]}$ ratio increases from 0.30 to 0.40 in response to a 20% reduction in $cytb_d$ activity (Fig. 5.6a). Alternatively, Fig. 5.6a shows that as O_2 levels increase, less oxidative power is required and the cell is able to produce more acetate, capitalising on the extra ATP generated by ACK resulting in a slight increase in the $\frac{[ATP]}{[ADP]}$ ratio.

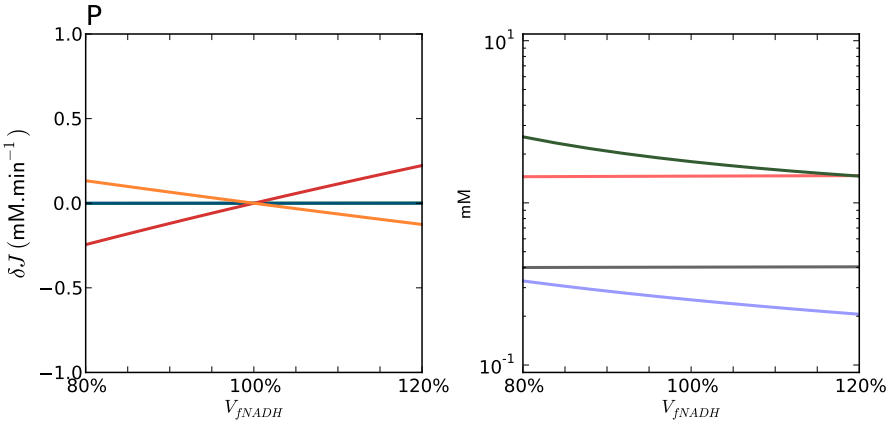
It was observed in Chapter 4 (Section 4.3.4), that the PFL rate determined under fully anaerobic conditions may in fact be too high to accurately represent the microaerobic state in which PFL is not fully expressed, and can be both denatured by residual O_2, and/or deactivated by its ancillary enzyme [132, 270]. To assess the effect of a reduced PFL rate, model P was reconfigured with a ten-fold reduction in PFL rate capacity ($V_{fPFL} = 0.41$), and the EIICB concentration was increased to produce the same steady state flux as in the original model. With the low-PFL model, the same simulation was performed as above in which $vNADH$ rate was varied by ±20 % Fig. 5.6b). Interestingly, in the low-PFL model, LDH assumes responsibility from ADH for balancing redox. As acetate flux decreases LDH flux increases proportionately and the $\frac{[NADH]}{[NAD^+]}$ ratio increases from 0.17 to 0.21 over a 20% reduction in $vNADH$ activity (compared to an increase of 0.25 to 0.33 in the high PFL model). This shifting of the redox burden to LDH has been observed under

microaerobic conditions in *pflA* mutants [260, 279]. *Ack* and *pta* mutants are unable to grow fermentatively on glucose, as are *adh* mutants [342, 343]; double mutants of both acetate and ethanol genes however, are able to ferment glucose in a homolactic fashion [342]. This observation suggests that flux through PFL is favoured over LDH under fermentative conditions as the current models demonstrate, but that LDH is able to compensate for lower PFL activity and/or an increased redox burden.

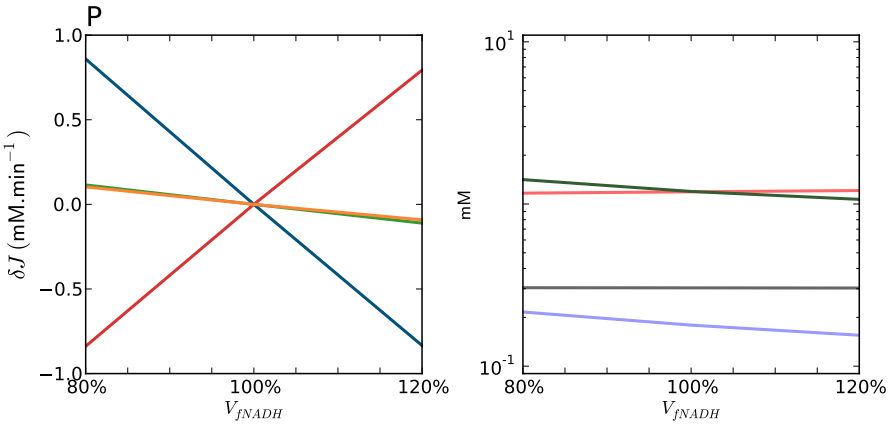
Thus, the apparent stoichiometric redundancy between lactate and acetate/ethanol fermentation affords the cell a degree of metabolic flexibility when facing varying O_2 levels and thus respiration levels, under microaerobic conditions. This relationship is also indicated by the negative flux control coefficient of $vLDH$ with respect to $vNADH$ in both the high and low PFL models: $C_{vNADH}^{JvLDH} = -1.56$ and $C_{vNADH}^{JvLDH} = -1.25$ respectively. The $cytb_d$ reaction affects LDH flux (but not PFL flux), presumably through the $\frac{[NADH]}{[NAD^+]}$ ratio which drives the redirection of flux from PFL and the ATP-generating acetate pathway to the oxidising lactate reaction. Though not included in the models constructed for this study, an allosteric response to high pyruvate concentrations increase the affinity of LDH for NADH [237], suggesting that at higher fluxes, pyruvate accumulation results in greater sensitivity of LDH to variations in $\frac{[NADH]}{[NAD^+]}$ and O_2. Additionally, lactate export has been shown to generate a proton motive force [136, 137] which may be an additional means of offsetting the loss of ATP production by the acetate pathway. Another effect produced by an increased $\frac{[NADH]}{[NAD^+]}$ ratio is an increase in FBP accumulation which may be a product of the inhibition of $vGAPDH$ by the change in $\frac{[NADH]}{[NAD^+]}$. The disequilibrium ratio of GAPDH is increased by the change in the $\frac{[NADH]}{[NAD^+]}$ ratio and thus the forward reaction of GAPDH requires more GAP to achieve a similar flux, resulting in more FBP and DHAP accumulated by the near-equilibrium ALD and TPI reactions. The effect of varying the $\frac{[NADH]}{[NAD^+]}$ ratio on GAPDH is a possible explanation for the accumulation of FBP in microaerobically grown *E. coli* and pfl mutants in general [279].

The ability of *E. coli* to respond to varying redox requirements by modulating the ratio of fermentation products allows the bacterium to metabolise substrates with different oxidation states [217]. Common alternative glycolytic substrates to glucose encountered by *E. coli* include sorbitol (glucitol) and mannitol which have a lower oxidation state than glucose (and thus produce more redox equivalents), and gluconate and glucuronate which have higher oxidation states that glucose [217][1]. Thus, a molecule of sorbitol or mannitol produces three reducing equivalents while glucose only produces two. As the only difference in cellular machinery in response to growth on mannitol, is an alternative mannitol-specific EIICBA component and the replacement of PGI with mannitol 1-phosphate 5-dehydrogenase, it is conceivable that the mechanisms presented above for maintaining redox balance apply in this situation too: under low O_2 tensions when PFL is fully active, balancing ethanol and acetate fermentation would compensate for the extra redox requirement; in the case of higher O_2 tensions in which PFL is either partially denatured or inactivated, LDH is capable of

[1] Böck and Sawers [217] calculate the oxidation state by assigning hydrogens a value of -0.5 and oxygens a value of +1.0.



(a)



(b)

Figure 5.6: Steady state response to changes in O_2 availability. Steady state variables were evaluated for model P (a), and (b) model P with a ten-fold reduction in PFL activity over a $cytb_d$ activity range of $\pm$ 20 % representing a response to decreased and increased O_2 availability respectively. The change in steady state flux (δJ) is plotted in the left hand block (J_{R9} ——, J_{vLDH} ——, J_{vACE} ——, J_{vETH} ——). Concentrations are plotted on the right ($\frac{[NADH]}{[NAD^+]}$ ——, $\frac{[ATP]}{[ADP]}$ ——, FBP ——, PEP ——). The right y-axis is log-scaled.

oxidising the extra reducing equivalents.

5.8 Conclusion

The control features of *E. coli* glycolysis under microaerobic conditions and the ability of the central carbon metabolic pathway in to respond to variations in O_2 and energy requirements has been explored by means of a number of *in silico* experiments. Elementary mode analysis, which uses only structural information, has elucidated the fundamental structural features of the fermentative glycolytic pathway. The two core fermentative routes to lactate and acetate/ethanol are identified as well as the aerobic acetogenic overflow fermentation route. In addition to these routes, a low ATP-generating flux route was identified which bypasses PK, suggesting a method whereby *pykA* mutants could maintain viability, as indeed they do [247]. Upon inclusion of the PCK reaction, a futile cycle is introduced which extends the flexibility of the fermentative pathways producing two new unique flux modes: one which is the futile cycle itself, with no net increase or decrease in metabolites as ATP-demand stoichiometrically compensates for the joint ATP-producing interaction of PPC and PCK—the simultaneous activity of PPC and PCK is known to be responsible for up to an 8 % loss of ATP in the WT (possibly to mobilise PO_4^{3-}) [122], however this involves cycling in the opposite direction to the current reported mode—and a high-energy yielding pathway which has been employed successfully as a target for metabolic engineering of a succinate producing strain of *E. coli* [335], displaying the utility of elementary mode analysis for informing synthetic biology projects.

A number of questions were posed in the beginning of this chapter and are responded to here:

1. *Is the "textbook" concept of a rate-limiting step in glycolysis valid for this model organism (usually PFK)?*

   In all models supply reactions controlled the flux, and the distribution of flux control is dominated by the glucose importing PTS pathway, and specifically by $R8$, the $IICB \cdot P \cdot EIIA$ association/dissociation reaction and less so by $R10$, the glucose transport step. The ATP demand makes almost no contribution to the control of flux. A consideration of the flux response coefficients for the concentrations of the various components of the PTS indicate that EIICB concentration exhibits the highest effect on flux. To this degree, the PTS represents a rate-limiting reaction pathway, and EIICB a flux controlling component with $R8$ and $R10$ (and less so $R7$ and $R3$) comprising the rate-limiting step/steps.

2. *Do ATP-demanding reactions external to glycolysis control flux through glycolysis in the microaerobic state (as shown for aerobic E. coli [101])?*

   The results of the *in silico* experiments with all models suggest unambiguously that under the

current experimental conditions the flux through glycolysis to ATP is controlled primarily by supply, which is contrary to the observations of Koebmann *et al.* [101] under aerobic conditions. If indeed the demand block exhibited a significantly lower elasticity (as in Koebmann *et al.* [101]), then the control would be shared between the blocks. However artificially increasing demand in aerobic cells using weak acids has also shown little or no effect on glucose uptake, and thus ATP supply [215], suggesting that it is quite possible for supply to control glycolytic flux in *E. coli*. It may be the case that EIICB concentrations, which are known to vary significantly *in vivo* [76, 77], may be poised at a branch point between supply and demand control of flux. EIICB concentrations are indeed close to this point in wild type cells [268]. An additional consequence of a lower demand elasticity is that the $\frac{[ATP]}{[ADP]}$ ratio would be free to vary over a broader range *in vivo*.

3. *Is the correlation between cellular free energy and PTS activity positive [118] or negative [101]?*

   The models suggest that there is indeed a positive correlation between the $\frac{[ATP]}{[ADP]}$ ratio and the rate of glucose uptake by the PTS as reported by Rohwer *et al.* [118]. This is possible by virtue of the competition of late glycolytic reactions for PEP, the phosphoryl donor used to import glucose into the cell.

4. *Do any particular enzymes represent key regulatory points in glycolysis, especially with regard to the cellular free energy level and redox balance?*

   Separating the supply block elasticity into partial response coefficients identifies PCK and PK as the key enzymes through which changes in the demand for ATP exert control over the flux of the supply block reactions, albeit a minor effect. The regulatory mechanism of the kinases appears to work by competing with the PTS for the common substrate PEP. If the demand block elasticity were as low as that reported by Koebmann *et al.* [101] then the effect of PCK and PK on the supply flux would be expected to be far more significant. As the demand block simply consists of a single mass-action equation, the supply block of the model simply exerts its control of demand flux in a linear mass action fashion by supplying more ATP. Primarily it is EIICB which controls the free energy state. PFL plays an important role in redox balancing and alterations in the activity of PFL result in a reconfiguration of ancillary reactions in response to changing redox demands.

5. *Are there structural features of glycolysis in E. coli which provide a unique degree of flexibility in response to redox demands?*

   It was shown that the ability of *E. coli* to maintain redox balance in response to changes

in O_2 availability under microaerobic conditions is mediated by a similar rearrangement of fermentation products as under anaerobiosis, as flux through the oxidative ethanol pathway and the non-oxidative acetate pathway are held in tension, and the flux ratio between these pathways is altered to generate a specific oxidative capacity. However, in the case of reduced PFL activity under microaerobiosis [132], LDH is able to perform the same redox balancing as ADH and remains in balance with acetate fermentation. This is primarily due to the saturation of the low activity ethanol pathway. The reason for the presence of LDH has long been discussed as it is stoichiometrically redundant in terms of redox balancing. This result suggests a role for the presence of lactate fermentation under varying O_2 availabilities.

6. *To what degree is E. coli able to respond metabolically to slight variations in levels of O_2, and how much of the O_2 adaptation requires a transcriptional response?*

   It is difficult to draw any rigorous conclusions in the absence of a thorough quantification of O_2 availability and respiratory activity similar to the study by Alexeeva *et al.* [132]. However, it does appear that the O_2-buffering design of the PEP-PYR node is capable of maintaining flux, the $\frac{[ATP]}{[ADP]}$ ratio, and the redox balance of the cell in at similar values in the absence of a transcriptional response, down to a $cytb_d$ activity of below 50 % in Model P (data not shown), which suggests a significant degree of flexibility mediated by the enzyme network alone.

7. *Is it possible for FBP to act as an internal glycolytic flux-sensor (as suggested by Kochanowski* et al. *[153])?*

   It has been shown that FBP does increase with glycolytic flux and is thus certainly capable of signalling flux changes. However, a key element of the flux-sensing theory posited by Kochanowski *et al.* [153] is that FBP increases linearly with glycolytic flux (this is made possible by the feed-forward stimulation of PK by FBP). Thus, whether FBP can in fact act as a flux sensor is, according to the current models, dependent on the degree of linearity required. However, model P does display a very slight degree of non-linearity, suggesting that this mechanism may indeed be plausible.

The results of this chapter show the power of the modelling approach for understanding the regulation of biological systems and suggest that the ability of *E. coli* to alter the ratio of mixed-acid fermentation products under microaerobic conditions provides a metabolic flexibility to the cell allowing it to respond to varying energy and redox demands, and giving it the ability to metabolise a host of possible substrates.

6 General discussion

Summary

A general discussion is presented in which the findings of the research presented in this dissertation are reviewed and their significance is discussed. A critique of the methodologies and results is offered. Finally, after suggestions are made for future research directions, the research context of this study is explored as this dissertation is grounded in a particular tradition offering a specific kind of solution to the problems faced by modern biology.

6.1 Synopsis and discussion

This dissertation has explored glycolysis and the fermentative pathways in *E. coli* under microaerobic conditions using a combination of *in vitro*, *in situ*, and *in vivo* experiments, resulting in a number of ODE models, suitable for probing *in silico*. The primary motivations of this study were manifold. A key motivation identifying *E. coli* as a desirable target for a modelling study was the relative scarcity of published kinetic models of central carbon metabolism in such an important model organism, especially when contrasted with the high number of constraint-based modelling studies [113] and the associated limitations of such techniques. Indeed, the necessity of a kinetic model of glycolysis in *E. coli* was confirmed by the fact that during the course of this project another kinetic model of central carbon metabolism in *E. coli* was produced by Peskov *et al.* [115], based on literature kinetic parameters and metabolomics data. A key difference between this study and the one by Peskov *et al.* [115] is that we sought to model *E. coli* glycolysis under conditions in which O_2 was not available in excess, and thereby to understand the nuances of the metabolic effects of O_2-limitation on the glycolytic and fermentative pathways and which (if any) unique metabolic strategies have been adopted by *E. coli*.

Additionally, the experimental richness of NMR technology, and its applicability to the field of systems biology was clearly perceived, though systems studies using NMR were sparse in the literature (a salient example of such a study is that by Mulquiney *et al.* [28, 344, 345] which explores 2,3-bisphosphoglycerate metabolism in erythrocytes using a modelling approach based on *in vivo* ^{13}C and ^{31}P NMR data) and often used unnaturally simplified mathematical representations (e.g. Exnowitz *et al.* [17]). Recent advances in the resolution attainable by NMR [17–19] are presenting NMR technology as an attractive tool for systems biology, precisely because NMR can be both non-invasive, yielding *in vitro* data, and because NMR technology is ideal for unravelling complexity by virtue of the number of metabolites that are simultaneously visible to the NMR spectrometer, and the broad scope of experimental designs available in the field of NMR. A brief synopsis of the ground covered by this project follows.

Chapter 2 presented NMRPy, a suite of open-source tools for NMR processing developed in the Python programming language. The goal in developing NMRPy was to provide biological and chemical scientists with a tool to easily and rapidly process NMR spectra and present refined data in an accessible visual format. Amongst the tools built into the NMRPy module are those necessary for the deconvolution of complex spectra, and array-processing capabilities, making NMRPy ideal for quantifying reaction time series, as performed in this study. Additionally, the built-in multi-threading capabilities and extensibility of the Python programming language make NMRPy ideal for fast batch-processing of large collections of data. NMRPy provided the data processing tools necessary for the current study and was used throughout.

Chapter 3 presented a new methodology for approximating enzyme kinetic parameters using NMR. It was shown that NMR is a superior technology for experimental observation of the behaviour of complex enzyme reaction networks, as many metabolites are quantified simultaneously,

including substrates, products, and allosteric effectors of the enzymes being studied. The efficacy of NMR reveals the limitations of traditional techniques such as coupled reaction studies using light spectrometry, in which changes in the concentration of an often significantly distant product (e.g. NAD[P]H) are observed, and accumulation of products or effectors, as well as unexpected competing reactions are not identifiable. Additionally, traditional techniques acquire only the initial data points of a reaction time series and approximate the initial rate of the reaction by linear fitting before significant product accumulation has taken place. In contrast time series methods of kinetic parameter approximation do not discard all the data after the first few linear points, but extract information from the whole time series. The result of these benefits is that far richer datasets are acquired per experiment and thus fewer experiments need to be performed in the characterisation of an enzymatic system.

The presented methodology was used to approximate kinetic parameters for the first two enzymes of glycolysis in *E. coli*: PGI and PFK. These two enzymes present an appropriate target of study for this methodology as they exhibit behaviours which are either invisible to traditional enzyme assay techniques or elucidated only by very cumbersome piecewise approaches: the PGI reaction is reversible and inhibited by its product, and PFK both binds F6P cooperatively and exhibits particularly complex allosteric regulatory behaviour toward several metabolites [186–189], being however most significantly inhibited by the late glycolytic intermediate PEP. It was shown that accumulating only five reaction time series for PGI, and 6 reaction time series for PFK was sufficient to fit PGI with a reversible Michaelis-Menten equation, and PFK with an irreversible Hill equation including substrate-binding cooperativity and inhibition by PEP. The fitted kinetic parameters were comparable with published values.

In Chapter 4 several models of glycolysis and the fermentation pathways under microaerobic conditions were developed using NMR data and published kinetic parameters. Due to the complexity of the lower half of glycolysis, and in particular the problem of not being able to quantify the degree to which each of several reactions consume shared moieties such as ATP/ADP, and $NADH/NAD^+$, the parameters of the lower reactions leading from DHAP/GAP to pyruvate were approximated by fitting thirteen aggregated NMR reaction time series with several glycolytic models. In addition to the core glycolytic pathway, the putative models included various combinations of ancillary reactions or pathways such as succinate fermentation, PEP-synthetase, adenylate kinase, and $cytb_d$-mediated NADH consumption. An information-theory approach was adopted to identify the best model structure, and selected models were expanded to cover mixed-acid fermentation pathways with the goal of emulating glucose uptake flux as determined by an NMR uptake assay. These models, designated O, P, Q and R, were evaluated by comparing their steady state properties to *in situ* experimental data with permeabilised cells, and on this basis model R was discarded. Models P and O outperformed the others.

Models were expanded to include the published PTS pathway by Rohwer *et al.* [68] and fermentation pathways, populated with literature parameters. Without fitting any of the PTS or

fermentation parameters the models achieved steady states at the published flux for the PTS model in isolation. To accurately emulate the *in vitro* environment, the feed-forward allosteric effect of FBP on PK was attenuated which permitted a slight accumulation of PEP, and the models subsequently attained the experimentally determined glucose uptake flux. Model steady state fluxes and metabolite concentrations were compared to NMR data detailing *in vivo* microaerobic fermentative flux distribution and internal metabolite accumulation, as well as published internal metabolite concentrations and fermentative fluxes. It was observed that all three models reproduced expected fermentative fluxes, and that model P—the simplest model produced, initially generated by fitting time series with only $cytb_d$ activity in addition to the core glycolytic pathway—was best able to emulate published internal microaerobic metabolite concentrations. The models generally predicted the accumulation of FBP, DHAP and 3PG, as reported in literature and shown to a degree by *in vivo* experimentation, however in this context Q and to a lesser degree O behaved more similarly to *pfl* mutants with a tendency to accumulate significantly higher concentrations of these metabolites.

Chapter 5 presented the results of a series of *in silico* analyses of these models, with the goal of elucidating structural and regulatory properties of the pathways under study which allow *E. coli* to respond to varying energy, and redox demands, including O_2 availability, and alternate substrates with redox states different to glucose. The first *in silico* experiment performed was an elementary mode analysis, identifying the sustainable fundamental fluxes attainable by the models. In addition to the basic modes covering recognised fermentative activity, an additional mode was identified which suggests how *E. coli* is able to bypass PK, and indeed does do so by rerouting flux in the case of *pykA* mutants [247]. Interestingly, upon the inclusion of PCK (as reported [122]), a futile cycle is formed which provides the cell with an extra degree of flexibility with regard to the $\frac{[ATP]}{[ADP]}$ ratio and PO_4^{3-}. This cycle results in the identification of a new flux mode in which significantly more succinate is produced and a higher energy output is achieved. This mode has been previously employed in metabolic engineering attempts towards a high-succinate producing strain of *E. coli* [335].

Metabolic control analysis was performed, and the control patterns of the models were compared. The models exhibited the same dominating element of the control distribution, that is, the PTS exhibited the highest control over glycolytic flux, and this control was centred on reaction $R8$, and to a lesser degree on $R10$. The high flux control coefficients, $R8$ and $R10$, identify EIICB, the glucose transporter, as a centre of flux control and a rate-limiting component. The reaction that was most responsive to alterations in the rate of other reactions—i.e. producing the highest flux control coefficients in general—was LDH. This control feature appears to implicate LDH as playing a key role in terms of the ability of *E. coli* to alter its flux distribution profile in response to energy and redox demands; a role which seems to be validated by altering these demands in the experiments described below.

To further understand the distribution of control in the model, a supply and demand analysis

was performed around the $\frac{[ATP]}{[ADP]}$ ratio, treating all of the glycolytic and fermentative reactions as a large ATP supply block, and a single reaction representing ATP demand. Unequivocally, control over glycolytic flux was dominated by the supply block, contrary to Koebmann *et al.* [101] and to a degree in agreement with Salmond *et al.* [215]. This result is partially dependent on the choice of rate law for the demand block, which in the current study was a simple mass-action equation. Assuming the demand block elasticity determined by Koebmann *et al.* [101] would mean that the supply and demand blocks shared control over glycolytic flux. A saturable rate law for the ATP demand reaction may indeed have produced a lower elasticity at sufficiently high ATP concentrations, however, the fact that ATP is part of a moiety conserved cycle implies that the demand reaction is inherently saturable. It has also been observed that the highly variable EIICB concentration may enable this flexibility by virtue of its closeness to an important point in aerobically grown wild type cells, beyond which control of glycolytic flux shifts to the supply block from the demand block [268].

By separating the supply block elasticity into partial response coefficients, it was shown that PCK and PK were the enzymes through which the demand block (via the $\frac{[ATP]}{[ADP]}$ ratio) exercised its minor control on the supply block flux. Ostensibly this was primarily mediated by the activation of these enzymes due to a lowering of the $\frac{[ATP]}{[ADP]}$ ratio, causing a competition between the kinases and the PTS for the common substrate PEP, and thus a reduction in glycolytic flux. A similar phenomenon has been observed in modelling glycolysis in *Saccharomyces cerevisae*, as increasing the ATP demand, lowered the $\frac{[ATP]}{[ADP]}$ ratio which reduced the positive effect of ATP on the early hexokinase step, thus reducing glycolytic flux [24]. This phenomenon would be expected to be significantly more important if the demand elasticity were lower.

A key ability of *E. coli* to regulate its redox response to changing O_2 levels under microaerobic conditions was elucidated in Chapter 5. At high PFL activities, resembling conditions that border on anaerobiosis, *E. coli* responds to changing redox requirements by a simple shift in the flux balance between acetate and ethanol fermentation [132]. The models emulated this behaviour. As PFL activity can be significantly reduced under microaerobic conditions [132, 270] due to transcriptional repression, O_2 denaturation, and the PFL inactivating enzyme, a version of model P was constructed with reduced PFL activity. It was shown that in the low-PFL model, lactate fermentation provided greater redox flexibility in response to changing O_2 levels. By holding the non-oxidative acetate flux and oxidative lactate flux in tension, *E. coli* is able to vary its oxidative capacity while maintaining stable $\frac{[NADH]}{[NAD^+]}$ and $\frac{[ATP]}{[ADP]}$ ratios and glycolytic flux under microaerobic conditions. This degree of flexibility is not observed when redox balancing is mediated solely by ACE and ADH, as ADH always remains close to saturation and expresses a low activity under microaerobic conditions [150]. It is assumed that this relationship applies to growth on more reduced substrates as well.

EIICB has been identified as a focal point of kinetic regulation in the central carbon metabolism of *E. coli*, and the presence at of LDH—which upon a cursory examination of the redox stoichiom-

etry of glycolysis and the fermentation pathways in *E. coli*, appears redundant—is thus presented as playing an important role in *E. coli* metabolism under microaerobic conditions, enabling the cell to respond rapidly and with greater flexibility to changing energy and redox demands.

6.2 Critique

A number of pertinent questions remain with regard to the methodologies and outcomes of this study and are explored in this section. As Chapter 3 was published before the final compilation of this dissertation and the paper has been reproduced verbatim, questions regarding the methodology and results presented therein are dealt within that chapter.

As an initial methodological consideration, it must be stated that one of the key limitations of NMR spectroscopy for studying complex reaction mixtures is the invisibility of metabolites not containing the observed nucleus. This can be overcome to a degree by monitoring two nucleotides instead of one [28]. However, ^{13}C NMR has the drawback of a low natural abundance of the NMR active nucleus, and ^{2}H can often be crowded or distorted by a large H_2O peak.

It may be asked whether a single model might not be able to replicate all of the scenarios presented in Chapters 4 and 5, especially if the topology is similar, simply by altering the V_{max} values. A model composed of a set of kinetic parameters not too dissimilar from reality may indeed be able to reproduce many desired steady states, however, the particular dynamic behaviour of a given system and its regulatory nuances—as exemplified by the control coefficient matrices presented in Fig. 5.2—can be expected to reside with the particular kinetic parameters chosen. This is certainly the case with regard to such complexities as binding cooperativity, product inhibition and allosteric effects. There is however a degree of redundancy in fitting both a binding constant and a V_{max}, in that in a typical saturable enzyme rate equation, increasing both the V_{max} and the binding constant together will produce a very similar linear region of the function around the half-saturation point. The implication of this phenomenon for parameter approximation is that there may be a tendency for these two parameters to be correlated during an iterative fitting procedure, and often a maximum parameter value has to be enforced to prevent the chosen fitting algorithm from allowing parameters to explode numerically. During the fitting procedures performed in this study, V_{max} values were constrained and not allowed to significantly exceed published values.

A reasonable consideration to bear in mind in any study employing experimental conditions different to the conditions about which inferences are to be made, is to what degree these environments are congruent. In this study the experimental *in vitro*, *in situ* and *in silico* environments were used to make inductions about the condition of the living cell *in vivo*. Indeed it has been shown that the differences between these environments in terms of ionic strength (especially with regard to the concentrations of potassium, PO_4^{3-} and sodium [257, 259, 346], and to a lesser degree calcium and sulphur [346]), and macromolecular crowding [75], can produce significantly different results *in vitro* than *in vivo*. A salient example is the reconditioning of the yeast glycolytic model

produced by Teusink *et al.* [24] with parameters determined under *in vivo*-like conditions [258]. The critique of *in vitro* assay buffers is generally that the ion or macromolecular concentrations are too low, and that they are often optimised for maximal activity not physiological levels of activity [346]. Sodium and potassium concentrations in the experiments presented in this study were closer to *in vivo* values by virtue of the fact that these are common counter ions for the buffer and metabolites included in the assay buffer (and typically NaOH or KOH was used to alter pH values of mixtures or buffers); phosphate was maintained at the reported physiological concentration and the allosteric effect of FBP on PK was attenuated accordingly [257]. Additionally, according to García-Contreras *et al.* [257], the core glycolytic activities in *E. coli* that are most affected by the difference between *in vitro* and *in vivo* assay buffers are the three adjacent TPI, GAPDH and PGK reactions. The GAPDH and PGK equilibrium constants do indeed exhibit a high degree of responsiveness to changes in ionic concentration, and some degree of accommodation has been made in this study for intracellular increases in ionic strength by altering the GAPDH equilibrium constant in all models according to the study by Cornell *et al.* [259]. During experimentation TPI would sporadically be inactivated in the course of collecting reaction time series. This could indicate the presence of an unknown ionic effect.

Regardless of these concerns, kinetic parameters are still ubiquitously determined under *in vitro* conditions in practice, imaginably due to the extra expense in terms of labour, time and money to compose an appropriate *in vivo*-like assay buffer. The *in situ* environment of the permeabilised cell has been treated conceptually in this study as representing essentially a compromise between the *in vitro* and *in vivo* environments. In the permeabilised cell, ionic concentrations are equalised with the assay buffer, as the buffer and substrates wash through the cell, but protein concentrations are essentially maintained at *in vivo* levels as large molecules are retained within the cell [253]. The ability of the models to reproduce the *in situ* data in Chapter 4, at the least suggests that macromolecular crowding, and possibly the differences in total intracellular concentration do not contribute too significantly to the difference between the *in vitro* and *in vivo* environments. The ability of the models to reproduce the observed *in vitro* fluxes and accumulation of metabolites, and published *in vivo* concentrations and fermentation distributions (even though the uptake flux and PFL rate were fitted) also suggests similarity between the *in vitro* models and the *in vivo* environment.

A final question demands attention: which model is correct? A simple response to this question is that no models are ever *sensu stricto* correct or true; all are "works of fiction" [347] grounded in a particular thought tradition offering an approximate and "fabled" [347] understanding of reality. Thus "the validation of a model is not that it is 'true' but that it generates good testable hypotheses relevant to important problems" [348], as with the structural, and free energy/redox experiments of Chapter 5. From this practical point of view and in the absence of further experimentation, model P appears to be superior to the other models developed, by virtue of its ability to reproduce the dynamic behaviour of *in situ* time courses, the observed *in vivo* fermentation fluxes, and published

steady state fluxes and internal metabolite concentrations. Additionally, model P was the simplest of the three models selected. It may be that a measure of over-fitting of model O to *in vitro* data allowed it to achieve the lowest Akaike score but at the cost of generalisability, producing a model that was less able to reproduce *in vivo* and *in situ* data when compared to model P. During the fitting of model Q, which had a high number of literature parameters and thus exhibited a lower flexibility toward the fitting algorithm, it may be that individual fitted parameters were pushed beyond reasonable values to achieve a fit.

6.3 Context and further research

The newly popular systems approach to biology grew out of cybernetics, the study of control mechanisms, a field which flourished in the middle of the twentieth century (e.g. Wiener [349]) before experiencing a period of relative quiet [350]. A motivating realisation which has fuelled the recent interest in systems biology is that reductive science has successfully identified the fundamental elements of biology, but it is unable to account for the emergent properties of complex systems [9], that is, properties exhibited solely by whole systems and not by any of the parts in isolation (e.g. regulatory feed-back loops) [351], and even more so as systems supervene upon and alter their components [352]. Reducing systems into elementary parts is able to account for the building blocks, but not by implication the complexity of their interrelationships. The modern paradigm of biological science which was inherited from Enlightenment thinkers is recognised as a necessary and highly productive transition period in history on the path to a systems view of nature [8]. A ground-breaking example of development beyond reductionist biology, is Robert Rosen's *Relational Biology* [353] and his emphasis on seeking principles which organise phenomena, rather than (or in addition to) the phenomena themselves [350]. This shifting of the focus of biological sciences and the fact that pure reductionism has run its course, demands a new degree of mathematical expertise on the part of the biologist [350]. This necessity demands reform of university biology courses in which exposure to key systems biology tools, such as mathematics and programming, is often kept to a minimum [354], or those in which the interdisciplinary nature of systems biology is not embraced [355]. Modelling provides a means of exploring the complex interrelationships of the fundamental elements of biology by encoding these components into mathematical structures. The holistic methodology adopted in this dissertation combines the riches of reductionist science with a systems perspective, providing room to anticipate the emergence of unpredictably complex properties of the system. Additionally the computational systems approach is fundamentally modular—models are often shared and can be recombined or used for new research—and thus lends itself to community development. It was encouraging to witness the ease with which the PTS model developed by Rohwer *et al.* [68] could be attached to the models developed thirteen years later in this study.

A number of system properties of glycolysis were discovered, including, the elementary flux

modes available to the system, the dominant control capacity of ATP demand on supply, the minor regulatory supply block access points that phosphoenolpyruvate carboxykinase and pyruvate kinase provide to the demand block, an affirmation of the possibility of FBP acting as an internal flux-sensor of glycolysis, and the various reorganisations of fermentation products (especially the balance between acetate and lactate fermentation) that provide the cell with an important degree of metabolic flexibility in response to fluctuating energy and redox requirements. In contrast, neither a purely reductionist (dividing glycolysis into component parts) nor a "top-down" (constraining a model to fit a particular phenotype) approach to modelling would necessarily have identified these key properties; the former sacrifices the general and emergent for the specific, the latter sacrifices the specific for the general, and in so doing is in danger of losing the particular nuances of the system under study, especially with regards to the dynamic responses of the system to external perturbations.

One of the key goals of this project was to produce scientific platforms for further research. Three such tools have been presented: NMRPy an NMR data processing tool kit with many applications in scientific research, the development of a spline-based methodology for approximating enzyme kinetic parameters from NMR reaction time series, and a model of glycolysis in *E. coli* under microaerobic conditions. A number of options exist to further validate the model building effort in this study and to explore the metabolism of *E. coli*. The primary distinction between the models produced, and some of the published observations under aerobic conditions, is a diametrically opposed pattern of flux control around the $\frac{[ATP]}{[ADP]}$ ratio. This difference can be verified experimentally under the microaerobic conditions employed in this study, by several means; the simplest possibly being a glucose uptake assay with weak acid perturbation of ATP demand (similar to the study by Salmond *et al.* [215]), and inhibition of glucose uptake to perturb ATP supply (e.g. with the non-metabolisable glucose analogue α-methylglucoside). Importantly, the experimental methodology presented is significantly less laborious when compared to traditional methods, and as such the process could informatively be repeated under aerobic and anaerobic conditions. Ultimately, *in vitro* time courses could be regenerated using an *in vivo*-like assay buffer, and the same models used in this study, fitted to these data. The comparison of the models generated under *in vivo*-like conditions and the results of the current study should prove informative.

It was argued in the beginning of this dissertation that the reason for certain microbes favouring low-yielding fermentative pathways even when abundant substrate is available was due to the faster rate of substrate-level phosphorylation compared to that of oxidative phosphorylation, and the lower investment in the protein machinery required. The kinetic model produced provides the platform to explore this behaviour quantitatively in comparison to aerobiosis. In terms of metabolic control, the complexities of the PTS and specifically the regulation of EIICB deserve further investigation, due to the high degree flux control exhibited by this component. Additionally, the relationship between pyruvate kinase and FBP—which possibly provides flux-sensing capabilities to the cell [153]—requires further elucidation. And finally, the precise relationship between O_2 availability, $cytb_d$

activity, and fermentation outputs should be quantified in a controlled chemostat environment, and compared to model predictions (e.g. short term reductions in O_2 under microaerobic conditions should decrease acetate production and increase lactate production).

Of general interest to identifying a "best practices" approach to systems biology model building is the question of to what degree literature parameter values are adequate for modelling purposes. In the current study, the best model (P) outperformed the model highly parameterised with literature values (Q). Exploring this difference will inevitably broach the topics of parameter identifiability in the context of time-course reaction data, and the transferability of experimentally determined parameters into alternate experimental contexts.

A kinetic model of glycolysis in *E. coli* can be a powerful tool, both in terms of its scientific explanatory power and as a platform for concept realisation, providing the ability to identify key regulatory points and to suggest focus areas for metabolic redesign in the context of synthetic biology applications, in fields as diverse and relevant as medicine, engineering and ecology. Indeed as the conclusions of this dissertation are derived from the study of a model organism, the regulatory principles presented may apply in general to the Enterobacteriaceae. The research presented in this dissertation is a positive confirmation of the value of the integrated experimental-computational systems approach to unravelling complex biological networks by reconstructing, rather than merely deconstructing them.

Bibliography

[1] Vincentelli, R. and Romier, C., 2013. Expression in *Escherichia coli*: becoming faster and more complex. *Curr. Opin. Struct. Biol.*, **23**(3):326.

[2] Lederberg, J. and Tatum, E. L., 1946. Gene recombination in *Escherichia coli*. *Nature*, **158**:558.

[3] Lederberg, E. M. and Lederberg, J., 1953. Genetic studies of lysogenicity in *Escherichia coli*. *Genetics*, **38**(1):51.

[4] Benzer, S., 1961. On the topography of the genetic fine structure. *Proc. Natl. Acad. Sci. USA*, **47**(3):403.

[5] Blount, Z. D., Borland, C. Z., and Lenski, R. E., 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **105**(23):7899.

[6] Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.*, 1997. The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**(5331):1453.

[7] McCloskey, D., Palsson, B. Ø., and Feist, A. M., 2013. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol.*, **9**(1):661.

[8] Woese, C. R., 2004. A new biology for a new century. *Microbiol. Mol. Biol. Rev.*, **68**(2):173.

[9] Sauer, U., Heinemann, M., and Zamboni, N., 2007. Getting closer to the whole picture. *Science*, **316**(5824):550.

[10] Snoep, J. L. and Westerhoff, H. V., 2005. From isolation to integration, a systems biology approach for building the silicon cell. In Alberghina, H. and Westerhoff, H. V. (eds.), *Systems Biology: Definitions and Perspectives*, pages 13–30. Springer.

[11] Kacser, H. and Burns, J., 1973. The control of flux. *Symp. Soc. Exp. Biol.*, **27**:65.

[12] Heinrich, R. and Rapoport, T. A., 1974. A linear steady-state treatment of enzymatic chains. *Eur. J. Biochem.*, **42**(1):97.

[13] Schuster, S., Dandekar, T., and Fell, D. A., 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.*, **17**(2):53.

[14] Hofmeyr, J.-H. S., 2001. Metabolic control analysis in a nutshell. *Proceedings of the 2nd International Conference on systems Biology*, pages 291–300.

[15] Hofmeyr, J. H. and Cornish-Bowden, A., 1991. Quantitative assessment of regulation in metabolic systems. *Eur. J. Biochem.*, **200**(1):223.

[16] Cornish-Bowden, A., 2013. The origins of enzyme kinetics. *FEBS Lett.*

[17] Exnowitz, F., Meyer, B., and Hackl, T., 2012. NMR for direct determination of K_{m} and $V_{\max}$ of enzyme reactions based on the Lambert W function-analysis of progress curves. *Biochim. Biophys. Acta*, **1824**(3):443.

[18] Ardenkjaer-Larsen, J. H., Fridlund, B., Gram, A., Hansson, G., Hansson, L., Lerche, M. H., Servin, R., Thaning, M., and Golman, K., 2003. Increase in signal-to-noise ratio of $>$ 10,000 times in liquid-state NMR. *Proc. Natl. Acad. Sci. USA*, **100**(18):10158.

[19] Meier, S., Jensen, P. R., and Duus, J. O., 2011. Real-time detection of central carbon metabolism in living *Escherichia coli* and its response to perturbations. *FEBS Lett.*, **585**(19):3133.

[20] Bruggeman, F. and Westerhoff, H., 2007. The nature of systems biology. *Trends Microbiol.*, **15**(1):45.

[21] Nicolas, T., 2009. Data mining, a tool for systems biology or a systems biology tool. *J. Comput. Sci. Syst. Biol.*, **2**(4):216.

[22] Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Assad-Garcia, N., Glass, J. I., and Covert, M. W., 2012. A whole-cell computational model predicts phenotype from genotype. *Cell*, **150**(2):389.

[23] Teusink, B., Walsh, M., van Dam, K., and Westerhoff, H., 1998. The danger of metabolic pathways with turbo design. *Trends Biochem. Sci.*, **23**(5):162.

[24] Teusink, B., Passarge, J., Reijenga, C., Esgalhado, E., van der Weijden, C., Schepper, M., Walsh, M., Bakker, B., van Dam, K., Westerhoff, H., *et al.*, 2000. Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.*, **267**(17):5313.

[25] Rohwer, J. M. and Botha, F. C., 2001. Analysis of sucrose accumulation in the sugar cane culm on the basis of *in vitro* kinetic data. *Biochem. J.*, **358**(Pt 2):437.

[26] Uys, L., Botha, F. C., Hofmeyr, J.-H. S., and Rohwer, J. M., 2007. Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. *Phytochemistry*, **68**(16-18):2375.

[27] du Preez, F. B., Conradie, R., Penkler, G. P., Holm, K., van Dooren, F. L. J., and Snoep, J. L., 2008. A comparative analysis of kinetic models of erythrocyte glycolysis. *J. Theor. Biol.*, **252**(3):488.

[28] Mulquiney, P. J., Bubb, W. A., and Kuchel, P. W., 1999. Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: *in vivo* kinetic characterization of 2,3-bisphosphoglycerate synthase/phosphatase using ^{13}C and ^{31}P NMR. *Biochem. J.*, **342**(3):567.

[29] Snoep, J., Penkler, G., and Rautenbach, M., 2009. Construction and validation of a detailed kinetic model for glycolysis in the asexual stage of *Plasmodium falciparum*; application for drug target identification. *N. Biotechnol.*, **25**:S354.

[30] Pillay, C. S., Hofmeyr, J.-H. S., and Rohwer, J. M., 2011. The logic of kinetic regulation in the thioredoxin system. *BMC Syst. Biol.*, **5**(1):15.

[31] Bakker, B. M., Michels, P. A., Opperdoes, F. R., and Westerhoff, H. V., 1997. Glycolysis in bloodstream form *Trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. *J. Biol. Chem.*, **272**(6):3207.

[32] Albert, M.-A., Haanstra, J. R., Hannaert, V., Roy, J. V., Opperdoes, F. R., Bakker, B. M., and Michels, P. A. M., 2005. Experimental and in silico analyses of glycolytic flux control in bloodstream form *Trypanosoma brucei*. *J. Biol. Chem.*, **280**(31):28306.

[33] Covert, M. W., ScHilling, C. H., Famili, I., Edwards, J. S., Goryanin, I. I., Sel'kov, E., and Palsson, B. Ø., 2001. Metabolic modeling of microbial strains *in silico*. *Trends Biochem. Sci.*, **26**(3):179.

[34] Varner, J. and Ramkrishna, D., 1999. Mathematical models of metabolic pathways. *Curr. Opin. Biotechnol.*, **10**(2):146.

[35] Schilling, C. H., Edwards, J. S., and Palsson, B. Ø., 1999. Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog.*, **15**(3):288.

[36] Kitano, H., 2002. Systems biology: a brief overview. *Science*, **295**(5560):1662.

[37] Tomita, M., 2001. Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol.*, **19**(6):205.

[38] Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J. C., and Hutchison, C. A., 1999. E-cell: software environment for whole-cell simulation. *Bioinformatics*, **15**(1):72.

[39] Snoep, J. L., 2005. The silicon cell initiative: working towards a detailed kinetic description at the cellular level. *Curr. Opin. Biotechnol.*, **16**(3):336.

[40] Rosen, R., 1991. *Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life*. Columbia University Press.

[41] Wang, E., 2010. *Cancer Systems Biology*. Chapman & Hall/CRC Mathematical & Computational Biology. Taylor & Francis.

[42] Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**(1):27.

[43] Karp, P., Riley, M., Paley, S., Pellegrini-Toole, A., and Krummenacker, M., 1999. ECOCYC: Encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res.*, **27**(1):55.

[44] Kurganov, B., 2007. *Systems biology: philosophical foundations*. Elsevier Science.

[45] Chen, X., Alonso, A. P., Allen, D. K., Reed, J. L., and Shachar-Hill, Y., 2011. Synergy between ^{13}C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in *E. coli*. *Metab. Eng.*, **13**(1):38.

[46] Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K., and Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.*, **79**(1).

[47] Palsson, B., 2000. The challenges of *in silico* biology. *Nat. Biotechnol.*, **18**(11):1147.

[48] Edwards, J. S., Covert, M., and Palsson, B., 2002. Metabolic modelling of microbes: the flux-balance approach. *Environ. Microbiol.*, **4**(3):133.

[49] Nanchen, A., Fuhrer, T., and Sauer, U., 2007. Determination of metabolic flux ratios from ^{13}C-experiments and gas chromatography-mass spectrometry data: protocol and principles. *Methods Mol. Biol.*, **358**:177.

[50] Peng, L., Arauzo-Bravo, M. J., and Shimizu, K., 2004. Metabolic flux analysis for a *ppc* mutant *Escherichia coli* based on ^{13}C-labelling experiments together with enzyme activity assays and intracellular metabolite measurements. *FEMS Microbiol. Lett.*, **235**(1):17.

[51] Pramanik, J. and Keasling, J. D., 1997. Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.*, **56**(4):398.

[52] Stephanopoulos, G., Aristidou, A., and Nielsen, J., 1998. *Metabolic Engineering: Principles and Methodologies*. Academic Press.

[53] Orth, J. D., Thiele, I., and Palsson, B. Ø., 2010. What is flux balance analysis? *Nat. Biotechnol.*, **28**(3):245.

[54] Varma, A. and Palsson, B. Ø., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.*, **60**(10):3724.

[55] Snitkin, E. and Segrè, D., 2008. Optimality criteria for the prediction of metabolic fluxes in yeast mutants. *Genome Inform.*, **20**:123.

[56] Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D., and Milo, R., 2011. The moderately efficient enzyme: evolutionary and physico-chemical trends shaping enzyme parameters. *Biochemistry (Mosc.)*, **50**(21):4402.

[57] Edwards, J. S., Ramakrishna, R., and Palsson, B. Ø., 2002. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol. Prog.*, **77**(1):27.

[58] Edwards, J. S. and Palsson, B. Ø., 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA*, **97**(10):5528.

[59] Kelk, S. M., Olivier, B. G., Stougie, L., and Bruggeman, F. J., 2012. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Sci. Rep.*, **2**:580.

[60] Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J. C., Montero, F., and Schuster, S., 1999. Metatool: for studying metabolic networks. *Bioinformatics*, **15**(3):251.

[61] Clarke, B. L., 1988. Stoichiometric network analysis. *Cell Biophys.*, **12**:237.

[62] Schilling, C., Letscher, D., and Palsson, B., 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.*, **203**(3):229.

[63] Liao, J. C., Hou, S. Y., and Chao, Y. P., 1996. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol. Prog.*, **52**(1):129.

[64] Schilling, C. H. and Palsson, B. Ø., 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. Theor. Biol.*, **203**(3):249.

[65] Papin, J. A., Price, N. D., Edwards, J. S., and Palsson, B. Ø., 2002. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J. Theor. Biol.*, **215**(1):67.

[66] Grivet, J. and Delort, A., 2009. NMR for microbiology: *In vivo* and *in situ* applications. *Pog. Nucl. Mag. Res. Sp.*, **54**(1):1.

[67] den Hollander, J. A., Brown, T. R., Ugurbil, K., and Shulman, R. G., 1979. ^{13}C nuclear magnetic resonance studies of anaerobic glycolysis in suspensions of yeast cells. *Proc. Natl. Acad. Sci. USA*, **76**(12):6096.

[68] Rohwer, J. M., Meadow, N. D., Roseman, S., Westerhoff, H. V., and Postma, P. W., 2000. Understanding glucose transport by the bacterial phosphoenolpyruvate:glycose phosphotransferase system on the basis of kinetic measurements *in vitro*. *J. Biol. Chem.*, **275**(45):34909.

[69] Postma, P. W., Lengeler, J. W., and Jacobson, G. R., 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.*, **57**(3):543.

[70] Deutscher, J., Francke, C., and Postma, P. W., 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.*, **70**(4):939.

[71] Prescott, L., Harley, J., and Klein, D., 1996. *Microbiology. 3rd Ed.* Mc Graw Hill. Newyork.

[72] Meadow, N. D., Fox, D. K., and Roseman, S., 1990. The bacterial phosphoenolpyruvate: glycose phosphotransferase system. *Annu. Rev. Biochem.*, **59**:497.

[73] Kholodenko, B. N. and Westerhoff, H. V., 1995. The macroworld versus the microworld of biochemical regulation and control. *Trends Biochem. Sci.*, **20**(2):52.

[74] Kholodenko, B. N., Cascante, M., and Westerhoff, H. V., 1995. Control theory of metabolic channelling. *Mol. Cell. Biochem.*, **143**(2):151.

[75] Rohwer, J. M., Postma, P. W., Kholodenko, B. N., and Westerhoff, H. V., 1998. Implications of macromolecular crowding for signal transduction and metabolite channeling. *Proc. Natl. Acad. Sci. USA*, **95**(18):10547.

[76] Stock, J. B., Waygood, E. B., Meadow, N., Postma, P., and Roseman, S., 1982. Sugar transport by the bacterial phosphotransferase system. the glucose receptors of the *Salmonella typhimurium* phosphotransferase system. *J. Biol. Chem.*, **257**(23):14543.

[77] Rephaeli, A. W. and Saier, M. H., 1980. Regulation of genes coding for enzyme constituents of the bacterial phosphotransferase system. *J. Bacteriol.*, **141**(2):658.

[78] Rohwer, J. M., Hanekom, A. J., Crous, C., Snoep, J. L., and Hofmeyr, J. H. S., 2006. Evaluation of a simplified generic bi-substrate rate equation for computational systems biology. *Syst. Biol. (Stevenage)*, **153**(5):338.

[79] Hanekom, A., 2006. *Generic kinetic equations for modelling multisubstrate reactions in computational systems biology*. Master's thesis, University of Stellenbosch. http://hdl.handle.net/10019.1/1659.

[80] Michaelis, L. and Menten, M., 1913. Die Kinetik der Invertinwirkung. *Biochem. Z.*, **49**:333.

[81] Henri, V., 1903. *Lois générales de l'action des diastases*. Librairie Scientifique A. Hermann.

[82] Briggs, G. E. and Haldane, J. B., 1925. A note on the kinetics of enzyme action. *Biochem. J.*, **19**(2):338.

[83] Hofmeyr, J. H. and Cornish-Bowden, A., 1997. The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models. *Comput. Appl. Biosci.*, **13**(4):377.

[84] Cornish-Bowden, A., 1995. *Fundamentals of Enzyme Kinetics*. Portland Press Ltd.

[85] Kurganov, B. I., Brookes, R., and Yakovlev, V., 1982. *Allosteric enzymes: kinetic behaviour*. J. Wiley.

[86] Hill, A., 1910. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.*, **40**(4).

[87] Hanekom, A. J., Hofmeyr, J. H. S., Snoep, J. L., and Rohwer, J. M., 2006. Experimental evidence for allosteric modifier saturation as predicted by the bi-substrate Hill equation. *Syst. Biol. (Stevenage)*, **153**(5):342.

[88] Rohwer, J. M., Hanekom, A. J., and Hofmeyr, J.-H. S., 2007. A universal rate equation for systems biology. *Experimental Standard Conditions of Enzyme Characterizations. Proceedings of the 2nd International Beilstein Workshop*, pages 175–187.

[89] Popova, S. V. and Sel'kov, E. E., 1975. Generalization of the model by Monod, Wyman and Changeux for the case of a reversible monosubtrate reaction SR,TP. *FEBS Lett.*, **53**(3):269.

[90] Boiteux, A., Markus, M., Plesser, T., Hess, B., and Malcovati, M., 1983. Analysis of progress curves. Interaction of pyruvate kinase from *Escherichia coli* with fructose 1, 6-bisphosphate and calcium ions. *Biochem. J.*, **211**(3):631.

[91] Liebermeister, W. and Klipp, E., 2006. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Model.*, **3**(1):41.

[92] Liebermeister, W. and Klipp, E., 2006. Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data. *Theor. Biol. Med. Model.*, **3**(1):42.

[93] Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H., and Herwig, R., 2011. *Systems biology*. Wiley. com.

[94] Yablonskii, G., Bykov, V., Gorban, A., and Elokhin, V., 1991. *Kinetic models of catalytic reactions*. Elsevier, Amsterdam.

[95] Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J.-H., Hunter, P. J., Juty, N. S., Kasberger, J. L., Kremling, A., Kummer, U., Novère, N. L., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J., and Forum, S. B. M. L., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**(4):524.

[96] Bray, T., Paoli, J., Sperberg-McQueen, C., Maler, E., and Yergeau, F., 1998. Extensible markup language (XML). *World Wide Web Consortium Recommendation REC-xml-19980210.*

[97] Besanko, D. and Braeutigam, R., 2010. *Microeconomics*. Wiley.

[98] Hofmeyr, J. H. S. and Cornish-Bowden, A., 2000. Regulating the cellular economy of supply and demand. *FEBS Lett.*, **476**(1-2):47.

[99] Hofmeyr, J. H., 1995. Metabolic regulation: a control analytic perspective. *J. Bioenerg. Biomembr.*, **27**(5):479.

[100] Hofmeyr, J. H., Olivier, B. G., and Rohwer, J. M., 2000. Putting the cart before the horse: Designing a metabolic system in order to understand it. In *Technological and Medical Implications of Metabolic Control Analysis*, chapter 34, page 299. Kluwer Academic Publishers.

[101] Koebmann, B. J., Westerhoff, H. V., Snoep, J. L., Nilsson, D., and Jensen, P. R., 2002. The glycolytic flux in *Escherichia coli* is controlled by the demand for ATP. *J. Bacteriol.*, **184**(14):3909.

[102] Keseler, I. M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., Johnson, D. A., Krummenacker, M., Nolan, L. M., Paley, S., Paulsen, I. T., Peralta-Gil, M., Santos-Zavaleta, A., Shearer, A. G., and Karp, P. D., 2009. ECOCYC: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res.*, **37**(Database issue):D464.

[103] Karp, P., Keseler, I., Shearer, A., Latendresse, M., Krummenacker, M., Paley, S., Paulsen, I., Collado-Vides, J., Gama-Castro, S., Peralta-Gil, M., *et al.*, 2007. Multidimensional annotation of the *Escherichia coli* K-12 genome. *Nucleic Acids Res.*, **35**(22):7577.

[104] Majewski, R. A. and Domach, M. M., 1990. Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnol. Prog.*, **35**(7):732.

[105] Pramanik, J. and Keasling, J. D., 1998. Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnol. Prog.*, **60**(2):230.

[106] Covert, M. W. and Palsson, B. Ø., 2002. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.*, **277**(31):28058.

[107] Reed, J. L., Vo, T. D., ScHilling, C. H., and Palsson, B. Ø., 2003. An expanded genome-scale model of *Escherichia coli* K-12. *Genome Biol.*, **4**(9):R54.

[108] Schaub, J., Mauch, K., and Reuss, M., 2007. Metabolic flux analysis in *Escherichia coli* by integrating isotopic dynamic and isotopic stationary ^{13}C labeling data. *Biotechnol. Bioeng.*, **99**(5):1170.

[109] Nöh, K., Grönke, K., Luo, B., Takors, R., Oldiges, M., and Wiechert, W., 2007. Metabolic flux analysis at ultra short time scale: isotopically non-stationary ^{13}C labeling experiments. *J. Biotechnol.*, **129**(2):249.

[110] van Rijsewijk, B. R. B. H., Nanchen, A., Nallet, S., Kleijn, R. J., and Sauer, U., 2011. Large-scale ^{13}C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Mol. Syst. Biol.*, **7**(1):477.

[111] Kremling, A. and Flockerzi, D., 2012. Structural analysis of a core model for carbohydrate uptake in *Escherichia coli*. *J. Theor. Biol.*, **303**:62.

[112] Doi, A., Fujita, S., Matsuno, H., Nagasaki, M., and Miyano, S., 2011. Constructing biological pathway models with hybrid functional petri nets. *Stud. Health Technol. Inform.*, **162**:92.

[113] Reed, J. L. and Palsson, B. Ø., 2003. Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *J. Bacteriol.*, **185**(9):2692.

[114] Rizzi, M., Baltes, M., Theobald, U., and Reuss, M., 1997. *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. mathematical model. *Biotechnol. Bioeng.*, **55**(4):592.

[115] Peskov, K., Mogilevskaya, E., and Demin, O., 2012. Kinetic modelling of central carbon metabolism in *Escherichia coli*. *FEBS J.*, **279**(18):3374.

[116] Buziol, S., Bashir, I., Baumeister, A., Claassen, W., Noisommit-Rizzi, N., Mailinger, W., and Reuss, M., 2002. New bioreactor-coupled rapid stopped-flow sampling technique for measurements of metabolite dynamics on a subsecond time scale. *Biotechnol. Prog.*, **80**(6):632.

[117] Theobald, U., Mailinger, W., Reuss, M., and Rizzi, M., 1993. *In vivo* analysis of glucose-induced fast changes in yeast adenine nucleotide pool applying a rapid sampling technique. *Anal. Biochem.*, **214**(1):31.

[118] Rohwer, J. M., Jensen, P. R., Shinohara, Y., Postma, P. W., and Westerhoff, H. V., 1996. Changes in the cellular energy state affect the activity of the bacterial phosphotransferase system. *Eur. J. Biochem.*, **235**(1-2):225.

[119] Rohwer, J. M., Bader, R., Westerhoff, H. V., and Postma, P. W., 1998. Limits to inducer exclusion: inhibition of the bacterial phosphotransferase system by glycerol kinase. *Mol. Microbiol.*, **29**(2):641.

[120] Cleland, W., 1963. The kinetics of enzyme-catalyzed reactions with two or more substrates or products: I. nomenclature and rate equations. *Biochimica et Biophysica Acta (BBA)-Specialized Section on Enzymological Subjects*, **67**:104.

[121] Monod, J., Wyman, J., and Changeux, J. P., 1965. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.*, **12**:88.

[122] Sauer, U. and Eikmanns, B. J., 2005. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol. Rev.*, **29**(4):765.

[123] Molenaar, D., van Berlo, R., de Ridder, D., and Teusink, B., 2009. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.*, **5**:323.

[124] Fischer, E. and Sauer, U., 2003. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *J. Biol. Chem.*, **278**(47):46446.

[125] Hoque, M. A., Ushiyama, H., Tomita, M., and Shimizu, K., 2005. Dynamic responses of the intracellular metabolite concentrations of the wild type and *pykA* mutant *Escherichia coli* against pulse addition of glucose or NH_3 under those limiting continuous cultures. *Biochemical Engineering Journal*, **26**(1):38.

[126] Yang, C., Hua, Q., Baba, T., Mori, H., and Shimizu, K., 2003. Analysis of *Escherichia coli* anaplerotic metabolism and its regulation mechanisms from the metabolic responses to altered dilution rates and phosphoenolpyruvate carboxykinase knockout. *Biotechnol. Bioeng.*, **84**(2):129.

[127] Siddiquee, K. A. Z., Arauzo-Bravo, M. J., and Shimizu, K., 2004. Metabolic flux analysis of pykf gene knockout *Escherichia coli* based on ^{13}C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. *Appl. Microbiol. Biotechnol.*, **63**(4):407.

[128] Wilson, M., 2004. *Microbial inhabitants of humans: their ecology and role in health and disease*. Cambridge University Press.

[129] Neidhardt, F. C. (ed.), 1996. *Escherichia coli and Salmonella: Cellular and Molecular Biology*. ASM Press.

[130] Rossmann, R., Sawers, G., and Böck, A., 1991. Mechanism of regulation of the formate-hydrogenlyase pathway by oxygen, nitrate, and pH: definition of the formate regulon. *Mol. Microbiol.*, **5**(11):2807.

[131] Holms, H., 1996. Flux analysis and control of the central metabolic pathways in *Escherichia coli*. *FEMS Microbiol. Rev.*, **19**(2):85.

[132] Alexeeva, S., De Kort, B., Sawers, G., Hellingwerf, K., and de Mattos, M., 2000. Effects of limited aeration and of the *arcab* system on intermediary pyruvate catabolism in *Escherichia coli*. *J. Bacteriol.*, **182**(17):4934.

[133] Hickman, J. and Ashwell, G., 1960. Uronic acid metabolism in bacteria. II. purification and properties of d-altronic acid and d-mannonic acid dehydrogenases in *Escherichia coli*. *J. Biol. Chem.*, **235**(6):1566.

[134] Wolfe, A., 2005. The acetate switch. *Microbiol. Mol. Biol. Rev.*, **69**(1):12.

[135] Sokatch, J. R. *et al.*, 1969. *Bacterial physiology and metabolism.* London (& New York): Academic Press.

[136] Michels, P. A., Michels, J. P., Boonstra, J., and Konings, W. N., 1979. Generation of an electrochemical proton gradient in bacteria by the excretion of metabolic end products. *FEMS Microbiol. Lett.*, **5**(5):357.

[137] Brink, B. and Konings, W. N., 1980. Generation of an electrochemical proton gradient by lactate efflux in membrane vesicles of *Escherichia coli*. *Eur. J. Biochem.*, **111**(1):59.

[138] Leif, H., Weidner, U., Berger, A., Spehr, V., Braun, M., van Heek, P., Friedrich, T., Ohnishi, T., and Weiss, H., 1993. *Escherichia coli* NADH dehydrogenase I, a minimal form of the mitochondrial complex I. *Biochem. Soc. Trans.*, **21**(4):998.

[139] Dancey, G. F. and Shapiro, B. M., 1976. The NADH dehydrogenase of the respiratory chain of *Escherichia coli*. II. kinetics of the purified enzyme and the effects of antibodies elicited against it on membrane-bound and free enzyme. *J. Biol. Chem.*, **251**(19):5921.

[140] D'mello, R., Hill, S., and Poole, R., 1996. The cytochrome bd quinol oxidase in *Escherichia coli* has an extremely high oxygen affinity and two oxygen-binding haems: implications for regulation of activity *in vivo* by oxygen inhibition. *Microbiology*, **142**(4):755.

[141] Puustinen, A., Finel, M., Haltia, T., Gennis, R., and Wikstrom, M., 1991. Properties of the two terminal oxidases of *Escherichia coli*. *Biochemistry (Mosc.)*, **30**(16):3936.

[142] Calhoun, M. W., Oden, K. L., Gennis, R. B., de Mattos, M. J., and Neijssel, O. M., 1993. Energetic efficiency of *Escherichia coli*: effects of mutations in components of the aerobic respiratory chain. *J. Bacteriol.*, **175**(10):3020.

[143] Gunsalus, R., Park, S., *et al.*, 1994. Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the *arcAB* and *fnr* regulons. *Res. Microbiol.*, **145**(5-6):437.

[144] Amarasingham, C. R. and Davis, B. D., 1965. Regulation of alpha-ketoglutarate dehydrogenase formation in *Escherichia coli*. *J. Biol. Chem.*, **240**(9):3664.

[145] Schuster, S., Pfeiffer, T., and Fell, D. A., 2008. Is maximization of molar yield in metabolic networks favoured by evolution? *J. Theor. Biol.*, **252**(3):497.

[146] van Hoek, M. J. A. and Merks, R. M. H., 2012. Redox balance is key to explaining full vs. partial switching to low-yield metabolism. *BMC Syst. Biol.*, **6**:22.

[147] Pfeiffer, T., Schuster, S., and Bonhoeffer, S., 2001. Cooperation and competition in the evolution of ATP-producing pathways. *Science*, **292**(5516):504.

[148] MacLean, R. C. and Gudelj, I., 2006. Resource competition and social conflict in experimental populations of yeast. *Nature*, **441**(7092):498.

[149] Piskur, J., Rozpedowska, E., Polakova, S., Merico, A., and Compagno, C., 2006. How did *Saccharomyces* evolve to become a good brewer? *Trends Genet.*, **22**(4):183.

[150] Peng, L. and Shimizu, K., 2003. Global metabolic regulation analysis for *Escherichia coli* K12 based on protein expression by 2-dimensional electrophoresis and enzyme activity measurement. *Appl. Microbiol. Biotechnol.*, **61**(2):163.

[151] Roberts, J. D., 2000. *ABCs of Ft-Nmr*. University Science Books.

[152] James, T., 1998. *Fundamentals of NMR*. Department of Pharmaceutical Chemistry University of California San Francisco, CA 94143-0446 USA.

[153] Kochanowski, K., Volkmer, B., Gerosa, L., van Rijsewijk, B. R. H., Schmidt, A., and Heinemann, M., 2013. Functioning of a metabolic flux sensor in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **110**(3):1130.

[154] Olivier, B. G., Rohwer, J. M., and Hofmeyr, J.-H. S., 2005. Modelling cellular systems with PySCeS. *Bioinformatics*, **21**(4):560.

[155] Molloy, J. C., 2011. The open knowledge foundation: open data means better science. *PLoS Biol.*, **9**(12):e1001195.

[156] Raymond, E. S., 2001. *The Cathedral and the Bazaar–Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly.

[157] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A., 1995. NMRpipe: a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR*, **6**(3):277.

[158] Laatikainen, R., Niemitz, M., Korhonen, S., Hassinen, T., and Venäläinen, T., 2011. Perch NMR software.

[159] van Beek, J., 2007. MATNMR: A flexible toolbox for processing, analyzing and visualizing magnetic resonance data in Matlab®. *J. Magn. Reson.*, **187**(1):19.

[160] Nilsson, M., 2009. The DOSY toolbox: A new tool for processing PFG NMR diffusion data. *J. Magn. Reson.*, **200**(2):296.

[161] Helmus, J. J. and Jaroniec, C. P., 2012. NMRglue. *http://code.google.com/p/{NMR}glue*.

[162] Hunter, J. D., 2007. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.*, **9**(3):90.

[163] Jones, E., Oliphant, T., Peterson, P., *et al.*, 2001. SciPy: Open source scientific tools for Python.

[164] Stefan, D., Cesare, F. D., Andrasescu, A., Popa, E., Lazariev, A., Vescovo, E., Strbak, O., Williams, S., Starcuk, Z., Cabanas, M., van Ormondt, D., and Graveron-Demilly, D., 2009. Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package. *Measurement Science and Technology*, **20**(10):104035.

[165] Fernando Pérez, B. E. G., 2007. IPython: A system for interactive scientific computing, computing in science and engineering. *Comput. Sci. Eng.*, **9**:21.

[166] Wasilewski, F., 2010. Pywavelets: Discrete wavelet transform in python.

[167] Cooley, J. W. and Tukey, J. W., 1965. An algorithm for the machine calculation of complex fourier series. *Math. Comput*, **19**:297.

[168] Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.*, **11**(2):431.

[169] Eicher, J. J., Snoep, J. L., and Rohwer, J. M., 2012. Determining enzyme kinetics for systems biology with nuclear magnetic resonance spectroscopy. *Metabolites*, **2**(4):818.

[170] Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., and Schomburg, D., 2011. BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.*, **39**(Database issue):D670.

[171] Krebs, O., Golebiewski, M., Kania, R., Mir, S., Saric, J., Weidemann, A., Wittig, U., and Rojas, I., 2007. SABIO-RK: A data warehouse for biochemical reactions and their kinetics. *J. Integr. Bioinf.*, **4**(1):49.

[172] Schnell, S., Chappell, M., Evans, N., and Roussel, M., 2006. The mechanism distinguishability problem in biochemical kinetics: The single-enzyme, single-substrate reaction as a case study. *C. R. Biol.*, **329**(1):51.

[173] Fernley, H. N., 1974. Statistical estimations in enzyme kinetics. The integrated Michaelis equation. *Eur. J. Biochem.*, **43**(2):377.

[174] Duggleby, R. G., 1995. Analysis of enzyme progress curves by nonlinear regression. *Methods Enzymol.*, **249**:61.

[175] Michaelis, L., Menten, M. L., Johnson, K. A., and Goody, R. S., 2011. The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry (Mosc.)*, **50**(39):8264.

[176] Schnell, S. and Mendoza, C., 2000. Time-dependent closed form solutions for fully competitive enzyme reactions. *Bull. Math. Biol.*, **62**(2):321.

[177] Goudar, C., Harris, S., McInerney, M., and Suflita, J., 2004. Progress curve analysis for enzyme and microbial kinetic reactions using explicit solutions based on the Lambert *W* function. *J. Microbiol. Methods*, **59**(3):317.

[178] Barshop, B. A., Wrenn, R. F., and Frieden, C., 1983. Analysis of numerical methods for computer simulation of kinetic processes: development of kinsim–a flexible, portable system. *Anal. Biochem.*, **130**(1):134.

[179] Orsi, B. A. and Tipton, K. F., 1979. Kinetic analysis of progress curves. *Methods Enzymol.*, **63**:159.

[180] Gibon, Y., Blaesing, O. E., Hannemann, J., Carillo, P., Höhne, M., Hendriks, J. H. M., Palacios, N., Cross, J., Selbig, J., and Stitt, M., 2004. A Robot-based platform to measure multiple enzyme activities in *Arabidopsis* using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. *Plant Cell.*, **16**(12):3304.

[181] Oldham, K. G., 2006. *Radiometric Methods of Enzyme Assay.* John Wiley & Sons, Inc.

[182] Greis, K. D., 2007. Mass spectrometry for enzyme assays and inhibitor screening: an emerging application in pharmaceutical research. *Mass Spectrom. Rev.*, **26**(3):324.

[183] Hua, Q., Yang, C., Baba, T., Mori, H., and Shimizu, K., 2003. Responses of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. *J. Bacteriol.*, **185**(24):7053.

[184] Schreyer, R. and Böck, A., 1980. Phosphoglucose isomerase from *Escherischia coli* K10: Purification, properties and formation under aerobic and anaerobic conditions. *Arch. Microbiol.*, **127**(3):289.

[185] Blangy, D., 1968. Phosphofructokinase from *E. coli*: Evidence for a tetrameric structure of the enzyme. *FEBS Lett.*, **2**(2):109.

[186] Kemerer, V. R., Griffin, C. C., and Brand, L., 1975. Phosphofructokinase from *Escherichia coli*. *Methods Enzymol.*, **42**:91.

[187] Blangy, D., Buc, H., and Monod, J., 1968. Kinetics of the allosteric interactions of phosphofructokinase from *Escherichia coli*. *J. Mol. Biol.*, **31**(1):13.

[188] Doelle, H. and Manderson, G., 1971. Comparative studies of fructose 1, 6-diphosphate aldolase from *Escherichia coli* 518 and *Lactobacillus casei var. rhamnosus* atcc 7469. *Antonie van Leeuwenhoek*, **37**(1):21.

[189] Deville-Bonne, D., Laine, R., and Garel, J., 1991. Substrate antagonism in the kinetic mechanism of *E. coli* phosphofructokinase-1. *FEBS Lett.*, **290**(1-2):173.

[190] Peskov, K., Goryanin, I., and Demin, O., 2008. Kinetic model of phosphofructokinase-1 from *Escherichia coli*. *J. Bioinform. Comput. Biol.*, **6**(4):843.

[191] Glonek, T., 1992. ^{31}P NMR of mg-ATP in dilute solutions: Complexation and exchange. *Int. J. Biochem.*, **24**(10):1533.

[192] Bock, J. L., 1980. The binding of metal ions to ATP: a proton and phosphorus NMR investigation of diamagnetic metal–ATP complexes. *J. Inorg. Biochem.*, **12**(2):119.

[193] Wurster, B. and Schneider, F., 1970. Kinetics of glucosephosphate isomerase (ec 5.3.1.9) from yeast *in vitro* and its application to flux calculations for the fermentation pathway of anaerobic yeast cells. *Hoppe Seylers Z. Physiol. Chem.*, **351**(8):961.

[194] Press, W., Vetterling, W., Flannery, B., and Teukolsky, S., 1993. *Numerical recipes in C: the art of scientific computing. 2*, volume 1. Cambridge university press.

[195] Gao, H., Chen, Y., and Leary, J., 2005. Kinetic measurements of phosphoglucose isomerase and phosphomannose isomerase by direct analysis of phosphorylated aldose-ketose isomers using tandem mass spectrometry. *Int. J. Mass Spectrom.*, **240**(3):291.

[196] Midelfort, C., Gupta, R., and Rose, I., 1976. Fructose 1, 6-bisphosphate: isomeric composition, kinetics, and substrate specificity for the aldolases. *Biochemistry (Mosc.)*, **15**(10):2178.

[197] Horecker, B., Tsolas, O., and Lai, C., 1972. 6. Aldolases. *The Enzymes*, **7**:213.

[198] Böhme, H. J., Schellenberger, W., and Hofmann, E., 1975. Microcalorimetric determination of thermochemical parameters of the phosphofructokinase reaction. *Acta Biol. Med. Ger.*, **34**(1):15.

[199] Frey, C. M. and Stuehr, J. E., 1972. Interactions of divalent metal ions with inorganic and nucleoside phosphates. I. Thermodynamics. *J. Am. Chem. Soc.*, **94**(25):8898.

[200] Frey, C. M., Banyasz, J. L., and Stuehr, J. E., 1972. Interactions of divalent metal ions with inorganic and nucleoside phosphates. II. Kinetics of magnesium(II) with $HP_3O_{10}{}^{4-}$, ATP, CTP, $HP_2O_7{}^{3-}$,ADP, and CDP. *J. Am. Chem. Soc.*, **94**(26):9198.

[201] Selwyn, M. J., 1965. A simple test for inactivation of an enzyme during assay. *Biochim. Biophys. Acta*, **105**(1):193.

[202] Ugurbil, K., Rottenberg, H., Glynn, P., and Shulman, R., 1978. ^{31}P nuclear magnetic resonance studies of bioenergetics and glycolysis in anaerobic *Escherichia coli* cells. *Proc. Natl. Acad. Sci. USA*, **75**(5):2244.

[203] Bradford, M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **72**(1):248.

[204] Song, D. and Jacques, N., 1997. Cell disruption of *Escherichia coli* by glass bead stirring for the recovery of recombinant proteins. *Anal. Biochem.*, **248**(2):300.

[205] Strohhäcker, J., Graaf, A., Schoberth, S., Wittig, R., and Sahm, H., 1993. ^{31}P nuclear magnetic resonance studies of ethanol inhibition in *Zymomonas mobilis*. *Arch. Microbiol.*, **159**(5):484.

[206] Williams, T., Kelley, C., and many others, 2011. Gnuplot 4.4: an interactive plotting program. *http://gnuplot.sourceforge.net/*.

[207] Harrington, B. and many others, 2012. Inkscape: An open source vector graphics editor. *http://www.inkscape.org/*.

[208] Hoffee, P. and Englesberg, E., 1962. Effect of metabolic activity of the glucose permase of bacterial cells. *Proc. Natl. Acad. Sci. USA*, **48**(10):1759.

[209] Hoffee, P., Englesberg, E., and Lamy, F., 1964. The glucose permease system in bacteria. *Biochimica et Biophysica Acta (BBA)-Specialized Section on Biophysical Subjects*, **79**(2):337.

[210] Hernandez-Asensio, M., Ramirez, J. M., and Del Campo, F. F., 1975. The control by respiration of the uptake of α-methyl glucoside in *Escherichia coli* K12. *Arch. Microbiol.*, **103**(1):155.

[211] del Campo, F. F., Hernández-Asensio, M., and Ramírez, J. M., 1975. Transport of α-methyl glucoside in mutants of *Escherichia coli* K12 deficient in Ca^{2+}, Mg^{2+}-activated adenosine triphosphatase. *Biochem. Biophys. Res. Commun.*, **63**(4):1099.

[212] Reider, E., Wagner, E. F., and Schweiger, M., 1979. Control of phosphoenolpyruvate-dependent phosphotransferase-mediated sugar transport in *Escherichia coli* by energization of the cell membrane. *Proc. Natl. Acad. Sci. USA*, **76**(11):5529.

[213] Robillard, G. and Konings, W., 1981. Physical mechanism for regulation of phosphoenolpyruvate-dependent glucose transport activity in *Escherichia coli*. *Biochemistry (Mosc.)*, **20**(17):5025.

[214] Klein, W. L. and Boyer, P. D., 1972. Energization of active transport by *Escherichia coli*. *J. Biol. Chem.*, **247**(22):7257.

[215] Salmond, C. V., Kroll, R. G., and Booth, I. R., 1984. The effect of food preservatives on pH homeostasis in *Escherichia coli*. *J. Gen. Microbiol.*, **130**(11):2845.

[216] Rohwer, J. M., 1997. *Interaction of Functional Units in Metabolism*. Ph.D. thesis, University of Amsterdam.

[217] Böck, A. and Sawers, G., 1996. Fermentation. In Neidhardt, F. C. (ed.), *Escherichia coli and Salmonella*, chapter 18, pages 262–282. ASM Press.

[218] Fong, S. S., Nanchen, A., Palsson, B. Ø., and Sauer, U., 2006. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *J. Biol. Chem.*, **281**(12):8024.

[219] Barabote, R. D. and Saier, M. H., 2005. Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiol. Mol. Biol. Rev.*, **69**(4):608.

[220] Malcovati, M. and Kornberg, H., 1969. Two types of pyruvate kinase in *Escherichia coli* K12. *Biochim. Biophys. Acta*, **178**(2):420.

[221] Waygood, E. and Sanwal, B., 1974. The control of pyruvate kinases of *Escherichia coli*. *J. Biol. Chem.*, **249**(1):265.

[222] Ogawa, T., Mori, H., Tomita, M., and Yoshino, M., 2007. Inhibitory effect of phosphoenolpyruvate on glycolytic enzymes in *Escherichia coli*. *Res. Microbiol.*, **158**(2):159.

[223] Krietsch, W., 1975. [92] triosephosphate isomerase from yeast. *Methods Enzymol.*, **41**:434.

[224] Markus, M., Plesser, T., Boiteux, A., Hess, B., and Malcovati, M., 1980. Analysis of progress curves. Rate law of pyruvate kinase type I from *Escherichia coli*. *Biochem. J.*, **189**(3):421.

[225] Doelle, H. W., 1975. ATP-sensitive and ATP-insensitive phosphofructokinase in *Escherichia coli* K-12. *Eur. J. Biochem.*, **50**(2):335.

[226] Ramseier, T. M., 1996. *Cra* and the control of carbon flux via metabolic pathways. *Res. Microbiol.*, **147**(6-7):489.

[227] Bianchi, L., Dorigo, M., Gambardella, L. M., and Gutjahr, W. J., 2009. A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing: an international journal*, **8**(2):239.

[228] Mitchell, M., 1998. *An Introduction to Genetic Algorithms*. A Bradford book. MIT Press.

[229] Storn, R. and Price, K., 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.*, **11**(4):341.

[230] Kennedy, J. and Eberhart, R., 1995. Particle swarm optimization. *Neural Networks, 1995. Proceedings., IEEE International Conference on*, **4**:1942.

[231] Shi, Y. and Eberhart, R., 1998. A modified particle swarm optimizer. *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence.*, pages 69–73.

[232] Srinivas, M. and Patnaik, L. M., 1994. Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE T. Syst. Man. Cyb.*, **24**(4):656.

[233] Harik, G. R., 1997. *Learning gene linkage to efficiently solve problems of bounded difficulty using genetic algorithms*. Ph.D. thesis, University of Michigan.

[234] Cooper, R. and Kornberg, H., 1967. The direct synthesis of phosphoenolpyruvate from pyruvate by *Escherichia coli*. *P. Roy. Soc. Lond. B Bio.*, **168**(1012):263.

[235] Matsushita, K., Ohnishi, T., and Kaback, H. R., 1987. NADH-ubiquinone oxidoreductases of the *Escherichia coli* aerobic respiratory chain. *Biochemistry (Mosc.)*, **26**(24):7732.

[236] Tarmy, E. and Kaplan, N., 1968. Chemical characterization of D-lactate dehydrogenase from *Escherichia coli* B. *J. Biol. Chem.*, **243**(10):2579.

[237] Tarmy, E. and Kaplan, N., 1968. Kinetics of *Escherichia coli* B D-lactate dehydrogenase and evidence for pyruvate-controlled change in conformation. *J. Biol. Chem.*, **243**(10):2587.

[238] Courtright, J. and Henning, U., 1970. Malate dehydrogenase mutants in *Escherichia coli* K-12. *J. Bacteriol.*, **102**(3):722.

[239] Pecher, A., Zinoni, F., Jatisatienr, C., Wirth, R., Hennecke, H., and Böck, A., 1983. On the redox control of synthesis of anaerobically induced enzymes in *Enterobacteriaceae*. *Arch. Microbiol.*, **136**(2):131.

[240] Merlin, C., Masters, M., McAteer, S., and Coulson, A., 2003. Why is carbonic anhydrase essential to *Escherichia coli*? *J. Bacteriol.*, **185**(21):6415.

[241] Housecroft, C. and Sharpe, A., 2001. *Inorganic Chemistry*. Prentice Hall, Pearson Education Limited, Harlow, UK.

[242] Carroll, J. J. and Mather, A. E., 1992. The system carbon dioxide-water and the Krichevsky-Kasarnovsky equation. *J. Solution Chem.*, **21**(7):607.

[243] Kai, Y., Matsumura, H., Inoue, T., Terada, K., Nagara, Y., Yoshinaga, T., Kihara, A., Tsumura, K., and Izui, K., 1999. Three-dimensional structure of phosphoenolpyruvate carboxylase: a proposed mechanism for allosteric inhibition. *Proc. Natl. Acad. Sci. USA*, **96**(3):823.

[244] Katsura, I., Taguchi, M., Morikawa, M., and Katsuki, H., 1981. Regulation of *Escherichia coli* phosphoenolpyruvate carboxylase by multiple effectors *in vivo*. II. Kinetic studies with a reaction system containing physiological concentrations of ligands. *J. Biochem. (Tokyo)*, **90**(5):1321.

[245] Fogarty, T. C., 1989. Varying the probability of mutation in the genetic algorithm. *Proceedings of the third international conference on Genetic algorithms*, pages 104–109.

[246] Back, T., 1992. Self-adaptation in genetic algorithms. *Proceedings of the First European Conference on Artificial Life*, pages 263–271.

[247] Emmerling, M., Dauner, M., Ponti, A., Fiaux, J., Hochuli, M., Szyperski, T., Wüthrich, K., Bailey, J., and Sauer, U., 2002. Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. *J. Bacteriol.*, **184**(1):152.

[248] Sauer, U., Lasko, D. R., Fiaux, J., Hochuli, M., Glaser, R., Szyperski, T., Wüthrich, K., and Bailey, J. E., 1999. Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J. Bacteriol.*, **181**(21):6679.

[249] Burnham, K. P. and Anderson, D. R., 2002. *Model selection and multi-model inference: a practical information-theoretic approach*. Springer Verlag.

[250] Zalta, E. N., Nodelman, U., Allen, C., and Perry, J., 2013. Stanford encyclopedia of philosophy. http://plato.stanford.edu.

[251] Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, **19**(6):716.

[252] Sugiura, N., 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Stat.-Theor. M.*, **7**(1):13.

[253] Cánovas, M., Torroglosa, T., and Iborra, J., 2005. Permeabilization of *Escherichia coli* cells in the biotransformation of trimethylammonium compounds into L-carnitine. *Enzyme Microb. Technol.*, **37**(3):300.

[254] Aragón, J. and Sols, A., 1991. Regulation of enzyme activity in the cell: effect of enzyme concentration. *FASEB J.*, **5**(14):2945.

[255] Aragón, J. J., Felíu, J. E., Frenkel, R. A., and Sols, A., 1980. Permeabilization of animal cells for kinetic studies of intracellular enzymes: *in situ* behavior of the glycolytic enzymes of erythrocytes. *Proc. Natl. Acad. Sci. USA*, **77**(11):6324.

[256] Miozzari, G. F., Niederberger, P., and Hütter, R., 1978. Permeabilization of microorganisms by Triton X-100. *Anal. Biochem.*, **90**(1):220.

[257] García-Contreras, R., Vos, P., Westerhoff, H. V., and Boogerd, F. C., 2012. Why *in vivo* may not equal *in vitro* - new effectors revealed by measurement of enzymatic activities under the same *in vivo*-like assay conditions. *FEBS J.*, **279**(22):4145.

[258] van Eunen, K., Kiewiet, J. A. L., Westerhoff, H. V., and Bakker, B. M., 2012. Testing biochemistry revisited: how *in vivo* metabolism can be understood from *in vitro* enzyme kinetics. *PLoS Comput. Biol.*, **8**(4):e1002483.

[259] Cornell, N. W., Leadbetter, M., and Veech, R. L., 1979. Effects of free magnesium concentration and ionic strength on equilibrium constants for the glyceraldehyde phosphate dehydrogenase and phosphoglycerate kinase reactions. *J. Biol. Chem.*, **254**(14):6522.

[260] Zhu, J. and Shimizu, K., 2005. Effect of a single-gene knockout on the metabolic regulation in *Escherichia coli* for D-lactate production under microaerobic condition. *Metab. Eng.*, **7**(2):104.

[261] Becker, S., Vlad, D., Schuster, S., Pfeiffer, P., and Unden, G., 1997. Regulatory O_2 tensions for the synthesis of fermentation products in *Escherichia coli* and relation to aerobic respiration. *Arch. Microbiol.*, **168**(4):290.

[262] Frey, A. D., Fiaux, J., Szyperski, T., Wüthrich, K., Bailey, J. E., and Kallio, P. T., 2001. Dissection of central carbon metabolism of hemoglobin-expressing *Escherichia coli* by ^{13}C nuclear magnetic resonance flux distribution analysis in microaerobic bioprocesses. *Appl. Environ. Microbiol.*, **67**(2):680.

[263] Millard, C. S., Chao, Y. P., Liao, J. C., and Donnelly, M. I., 1996. Enhanced production of succinic acid by overexpression of phosphoenolpyruvate carboxylase in *Escherichia coli*. *Appl. Environ. Microbiol.*, **62**(5):1808.

[264] Lolkema, J. S., ten Hoeve-Duurkens, R., and Robillard, G. T., 1993. Steady state kinetics of mannitol phosphorylation catalyzed by enzyme IImtl of the *Escherichia coli* phosphoenolpyruvate-dependent phosphotransferase system. *J. Biol. Chem.*, **268**(24):17844.

[265] Lolkema, J. S., Kuiper, H., Ten Hoeve-Duurkens, R. H., and Robillard, G. T., 1993. Mannitol-specific enzyme II of the phosphoenolpyruvate-dependent phosphotransferase system of *Escherichia coli*: physical size of enzyme IImtl and its domains IIBA and IIC in the active state. *Biochemistry (Mosc.)*, **32**(6):1396.

[266] Boer, H., ten Hoeve-Duurkens, R., Schuurman-Wolters, G. K., Dijkstra, A., and Robillard, G. T., 1994. Expression, purification, and kinetic characterization of the mannitol transport domain of the phosphoenolpyruvate-dependent mannitol phosphotransferase system of *Escherichia coli*. kinetic evidence that the *E. coli* mannitol transport protein is a functional dimer. *J. Biol. Chem.*, **269**(27):17863.

[267] Roossien, F., Blaauw, M., and Robillard, G. T., 1984. Kinetics and subunit interaction of the mannitol-specific enzyme II of the *Escherichia coli* phosphotransferase system. *Biochemistry*, **23**(21):4934.

[268] Ruyter, G. J., Postma, P. W., and van Dam, K., 1991. Control of glucose metabolism by enzyme IIglc of the phosphoenolpyruvate-dependent phosphotransferase system in *Escherichia coli*. *J. Bacteriol.*, **173**(19):6184.

[269] Becker, A., Fritz-Wolf, K., Kabsch, W., Knappe, J., Schultz, S., and Wagner, A. V., 1999. Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nat. Struct. Mol. Biol.*, **6**(10):969.

[270] Conradt, H., Hohmann-Berger, M., Hohmann, H.-P., Blaschkowski, H. P., and Knappe, J., 1984. Pyruvate formate-lyase (inactive form) and pyruvate formate-lyase activating enzyme of *Escherichia coli*: Isolation and structural properties. *Arch. Biochem. Biophys.*, **228**(1):133.

[271] Kessler, D., Herth, W., and Knappe, J., 1992. Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of *Escherichia coli*. *J. Biol. Chem.*, **267**(25):18073.

[272] Nnyepi, M. R., Peng, Y., and Broderick, J. B., 2007. Inactivation of *E. coli* pyruvate formate-lyase: Role of AdhE and small molecules. *Arch. Biochem. Biophys.*, **459**(1):1.

[273] Zimmerman, S. B. and Trach, S. O., 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.*, **222**(3):599.

[274] Schultz, S. G. and Solomon, A., 1961. Cation transport in *Escherichia coli* I. Intracellular Na and K concentrations and net cation movement. *J. Gen. Physiol.*, **45**(2):355.

[275] Rosenberg, H., Russell, L. M., Jacomb, P. A., and Chegwidden, K., 1982. Phosphate exchange in the pit transport system in *Escherichia coli*. *J. Bacteriol.*, **149**(1):123.

[276] Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Dien, S. J. V., and Rabinowitz, J. D., 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.*, **5**(8):593.

[277] Nakanishi-Matsui, M., Sekiya, M., Nakamoto, R. K., and Futai, M., 2010. The mechanism of rotating proton pumping ATPases. *Biochim. Biophys. Acta*, **1797**(8):1343.

[278] Danchin, A., Dondon, L., and Daniel, J., 1984. Metabolic alterations mediated by 2-ketobutyrate in *Escherichia coli* K12. *Mol. Gen. Genet.*, **193**(3):473.

[279] Zhu, J. and Shimizu, K., 2004. The effect of pfl gene knockout on the metabolism for optically pure D-lactate production by *Escherichia coli*. *Appl. Microbiol. Biotechnol.*, **64**(3):367.

[280] Albe, K. R., Butler, M. H., and Wright, B. E., 1990. Cellular concentrations of enzymes and their substrates. *J. Theor. Biol.*, **143**(2):163.

[281] Lohmeier-Vogel, E. M., Ung, S., and Turner, R. J., 2004. *In vivo* ^{31}P nuclear magnetic resonance investigation of tellurite toxicity in *Escherichia coli*. *Appl. Environ. Microbiol.*, **70**(12):7342.

[282] Schneider, D. and Gourse, R., 2003. Changes in the concentrations of guanosine 5'-diphosphate 3'-diphosphate and the initiating nucleoside triphosphate account for inhibition of rRNA transcription in fructose-1, 6-diphosphate aldolase (fda) mutants. *J. Bacteriol.*, **185**(20):6192.

[283] Baldwin, S., Perham, R., and Stribling, D., 1978. Purification and characterization of the class-II D-fructose 1, 6-bisphosphate aldolase from *Escherichia coli* (Crookes' strain). *Biochem. J.*, **169**(3):633.

[284] Kobes, R. D., Simpson, R. T., Vallee, R. L., and Rutter, W. J., 1969. A functional role of metal ions in a class II aldolase. *Biochemistry (Mosc.)*, **8**(2):585.

[285] Thomas, A. D., Doelle, H. W., Westwood, A. W., and Gordon, G. L., 1972. Effect of oxygen on several enzymes involved in the aerobic and anaerobic utilization of glucose in *Escherichia coli*. *J. Bacteriol.*, **112**(3):1099.

[286] Utter, M. F. and Werkman, C. H., 1941. Occurrence of the aldolase and isomerase equilibria in bacterial metabolism. *J. Bacteriol.*, **42**(5):665.

[287] Zgiby, S. M., Thomson, G. J., Qamar, S., and Berry, A., 2000. Exploring substrate binding and discrimination in fructose1, 6-bisphosphate and tagatose 1,6-bisphosphate aldolases. *Eur. J. Biochem.*, **267**(6):1858.

[288] Babul, J., Clifton, D., Kretschmer, M., and Fraenkel, D., 1993. Glucose metabolism in *Escherichia coli* and the effect of increased amount of aldolase. *Biochemistry (Mosc.)*, **32**(17):4685.

[289] Cayley, S., Lewis, B. A., Guttman, H. J., and Record, M. T., 1991. Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity. Implications for protein-DNA interactions *in vivo*. *J. Mol. Biol.*, **222**(2):281.

[290] Veech, R. L., Raijman, L., Dalziel, K., and Krebs, H. A., 1969. Disequilibrium in the triose phosphate isomerase system in rat liver. *Biochem. J.*, **115**(4):837.

[291] Segel, I. H., 1976. *Biochemical calculations*. Wiley New York.

[292] Li, X., Wu, F., Qi, F., and Beard, D. A., 2011. A database of thermodynamic properties of the reactions of glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway. *Database*, **2011**.

[293] Seta, F. D., Boschi-Muller, S., Vignais, M., and Branlant, G., 1997. Characterization of *Escherichia coli* strains with *gapA* and *gapB* genes deleted. *J. Bacteriol.*, **179**(16):5218.

[294] Boschi-Muller, S., Azza, S., Pollastro, D., Corbier, C., and Branlant, G., 1997. Comparative enzymatic properties of GapB-encoded erythrose-4-phosphate dehydrogenase of *Escherichia coli* and phosphorylating glyceraldehyde-3-phosphate dehydrogenase. *J. Biol. Chem.*, **272**(24):15106.

[295] Eyschen, J., Vitoux, B., Marraud, M., Cung, M., and Branlant, G., 1999. Engineered glycolytic glyceraldehyde-3-phosphate dehydrogenase binds the anticonformation of NAD^+ nicotinamide but does not experience a-specific hydride transfer. *Arch. Biochem. Biophys.*, **364**(2):219.

[296] Byers, L. D., She, H. S., and Alayoff, A., 1979. Interaction of phosphate analogues with glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry (Mosc.)*, **18**(12):2471.

[297] Nellemann, L. J., Holm, F., Atlung, T., and Hansen, F. G., 1989. Cloning and characterization of the *Escherichia coli* phosphoglycerate kinase (*pgk*) gene. *Gene*, **77**(1):185.

[298] Fifis, T. and Scopes, R. K., 1978. Purification of 3-phosphoglycerate kinase from diverse sources by affinity elution chromatography. *Biochem. J.*, **175**(1):311.

[299] Fraser, H., Kvaratskhelia, M., and White, M., 1999. The two analogous phosphoglycerate mutases of *Escherichia coli*. *FEBS Lett.*, **455**(3):344.

[300] Foster, J., Davis, P., Raverdy, S., Sibley, M., Raleigh, E., Kumar, S., and Carlow, C., 2010. Evolution of bacterial phosphoglycerate mutases: non-homologous isofunctional enzymes undergoing gene losses, gains and lateral transfers. *PLoS One*, **5**(10):e13576.

[301] Clarke, J. B., Birch, M., and Britton, H. G., 1974. The equilibrium constant of the phosphoglyceromutase reaction. *Biochem. J.*, **139**(3):491.

[302] Wold, F. and Ballou, C., 1957. Studies on the enzyme enolase. I. Equilibrium studies. *J. Biol. Chem.*, **227**(1):301.

[303] Spring, T. and Wold, F., 1971. The purification and characterization of *Escherichia coli* enolase. *J. Biol. Chem.*, **246**(22):6797.

[304] Speranza, M. L., Valentini, G., Iadarola, P., Stoppini, M., Malcovati, M., and Ferri, G., 1989. Primary structure of three peptides at the catalytic and allosteric sites of the fructose-1,6-bisphosphate-activated pyruvate kinase from *Escherichia coli*. *Biol. Chem. Hoppe. Seyler*, **370**(3):211.

[305] Dobson, G., Hitchins, S., and Teague, W., 2002. Thermodynamics of the pyruvate kinase reaction and the reversal of glycolysis in heart and skeletal muscle. *J. Biol. Chem.*, **277**(30):27176.

[306] Valentini, G., Chiarelli, L., Fortin, R., Speranza, M. L., Galizzi, A., and Mattevi, A., 2000. The allosteric regulation of pyruvate kinase. *J. Biol. Chem.*, **275**(24):18145.

[307] Berman, K. M. and Cohn, M., 1970. Phosphoenolpyruvate synthetase. Partial reactions studied with adenosine triphosphate analogues and the inorganic phosphate-H2 18o exchange reaction. *J. Biol. Chem.*, **245**(20):5319.

[308] Nowak, T. and Mildvan, A. S., 1970. Stereoselective interactions of phosphoenolpyruvate analogues with phosphoenolpyruvate-utilizing enzymes. *J. Biol. Chem.*, **245**(22):6057.

[309] Sakuraba, H., Utsumi, E., Kujo, C., and Ohshima, T., 1999. An AMP-dependent (ATP-forming) kinase in the hyperthermophilic archaeon *Pyrococcus furiosus*: characterization and novel physiological role. *Arch. Biochem. Biophys.*, **364**(1):125.

[310] Bârzu, O. and Michelson, S., 1983. Simple and fast purification of *Escherichia coli* adenylate kinase. *FEBS Lett.*, **153**(2):280.

[311] Langer, R. S., Gardner, C. R., Hamilton, B. K., and Colton, C. K., 1977. Enzymatic regeneration of ATP: II. Equilibrium studies with acetate kinase and adenylate kinase. *AlChE J.*, **23**(1):1.

[312] Futai, M. and Kimura, H., 1977. Inducible membrane-bound l-lactate dehydrogenase from *Escherichia coli*. purification and properties. *J. Biol. Chem.*, **252**(16):5820.

[313] Mat-Jan, F., Alam, K. Y., and Clark, D. P., 1989. Mutants of *Escherichia coli* deficient in the fermentative lactate dehydrogenase. *J. Bacteriol.*, **171**(1):342.

[314] Meinardus-Hager, G. and Gäde, G., 1986. The separate function of d-lactate-, octopine-, and alanopine dehydrogenases in the foot muscle of the jumping cockle *Cardium tuberculatum* during anaerobiosis. *J. Comp. Physiol., B*, **156**(6):873.

[315] Knappe, J. and Blaschkowski, H. P., 1975. Pyruvate formate-lyase from escherischia coli and its activation system. *Methods Enzymol.*, **41**:508.

[316] Fox, D. K. and Roseman, S., 1986. Isolation and characterization of homogeneous acetate kinase from *Salmonella typhimurium* and *Escherichia coli*. *J. Biol. Chem.*, **261**(29):13487.

[317] Shimizu, M., Suzuki, T., Kameda, K.-Y., and Abiko, Y., 1969. Phosphotransacetylase of *Escherichia coli* b, purification and properties. *Biochimica et Biophysica Acta (BBA)-Enzymology*, **191**(3):550.

[318] Brinsmade, S. R. and Escalante-Semerena, J. C., 2007. *In vivo* and *in vitro* analyses of single-amino acid variants of the *Salmonella enterica* phosphotransacetylase enzyme provide insights into the function of its n-terminal domain. *J. Biol. Chem.*, **282**(17):12629.

[319] Shone, C. C. and Fromm, H. J., 1981. Steady-state and pre-steady-state kinetics of coenzyme a-linked aldehyde dehydrogenase from *Escherichia coli*. *Biochemistry (Mosc.)*, **20**(26):7494.

[320] Iverson, T. M., Luna-Chavez, C., Cecchini, G., and Rees, D. C., 1999. Structure of the *Escherichia coli* fumarate reductase respiratory complex. *Science*, **284**(5422):1961.

[321] Corwin, L. M. and Fanning, G. R., 1968. Studies of parameters affecting the allosteric nature of phosphoenolpyruvate carboxylase of *Escherichia coli*. *J. Biol. Chem.*, **243**(12):3517.

[322] Jetten, M. S. and Sinskey, A. J., 1995. Purification and properties of oxaloacetate decarboxylase fromcorynebacterium glutamicum. *Antonie van Leeuwenhoek*, **67**(2):221.

[323] Tchen, T., Loewus, F. A., and Vennesland, B., 1955. The mechanism of enzymatic carbon dioxide fixation into oxalacetate. *J. Biol. Chem.*, **213**(2):547.

[324] Muslin, E. H., Li, D., Stevens, F. J., Donnelly, M., Schiffer, M., and Anderson, L. E., 1995. Engineering a domain-locking disulfide into a bacterial malate dehydrogenase produces a redox-sensitive enzyme. *Biophys. J.*, **68**(6):2218.

[325] Guynn, R. W., Gelberg, H. J., and Veech, R. L., 1973. Equilibrium constants of the malate dehydrogenase, citrate synthase, citrate lyase, and acetyl coenzyme a hydrolysis reactions under physiological conditions. *J. Biol. Chem.*, **248**(20):6957.

[326] Woods, S. A., Schwartzbach, S. D., and Guest, J. R., 1988. Two biochemically distinct classes of fumarase in *Escherichia coli*. *Biochim. Biophys. Acta*, **954**(1):14.

[327] Gajewski, E., Goldberg, R., and Steckler, D., 1985. Thermodynamics of the conversion of fumarate to l-(-)-malate. *Biophys. Chem.*, **22**(3):187.

[328] Lemire, B. D., Robinson, J. J., and Weiner, J. H., 1982. Identification of membrane anchor polypeptides of *Escherichia coli* fumarate reductase. *J. Bacteriol.*, **152**(3):1126.

[329] Utter, M. F. and Kolenbrander, H. M., 1972. 4. Formation of oxalacetate by CO_2 fixation on phosphoenolpyruvate. *The enzymes*, **6**:117.

[330] Goldie, A. H. and Sanwal, B. D., 1980. Genetic and physiological characterization of *Escherichia coli* mutants deficient in phosphoenolpyruvate carboxykinase activity. *J. Bacteriol.*, **141**(3):1115.

[331] Krebs, A. and Bridger, W. A., 1980. The kinetic properties of phosphoenolpyruvate carboxykinase of *Escherichia coli*. *Can. J. Biochem.*, **58**(4):309.

[332] Jomain-Baum, M. and Schramm, V., 1978. Kinetic mechanism of phosphoenolpyruvate carboxykinase (GTP) from rat liver cytosol. product inhibition, isotope exchange at equilibrium, and partial reactions. *J. Biol. Chem.*, **253**(10):3648.

[333] Westermark, P., Hellgren-Kotaleski, J., and Lansner, A., 2004. Derivation of a reversible hill equation with modifiers affecting catalytic properties. *WSEAS Transactions on Biology and Medicine*, **1**:91.

[334] Wood, H. G., Davis, J. J., and Lochmüller, H., 1966. The equilibria of reactions catalyzed by carboxytransphosphorylase, carboxykinase, and pyruvate carboxylase and the synthesis of phosphoendolpyruvate. *J. Biol. Chem.*, **241**(23):5692.

[335] Kim, P., Laivenieks, M., Vieille, C., and Zeikus, J. G., 2004. Effect of overexpression of actinobacillus succinogenes phosphoenolpyruvate carboxykinase on succinate production in *Escherichia coli*. *Appl. Environ. Microbiol.*, **70**(2):1238.

[336] Zill, D. G. and Wright, W. S., 2009. *Advanced engineering mathematics*. Jones & Bartlett Publishers.

[337] Arrowsmith, D. K., 1990. *An introduction to dynamical systems*. Cambridge University Press.

[338] van der Vlag, J., van't Hof, R., van Dam, K., and Postma, P. W., 1995. Control of glucose metabolism by the enzyme of the glucose phosphotransferase system in *Salmonella typhimirium*. *Eur. J. Biochem.*, **230**.

[339] Rohwer, J. M. and Hofmeyr, J.-H. S., 2008. Identifying and characterising regulatory metabolites with generalised supply-demand analysis. *J. Theor. Biol.*, **252**(3):546.

[340] Kotte, O., Zaugg, J. B., and Heinemann, M., 2010. Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol. Syst. Biol.*, **6**(1):355.

[341] Saier, M. H. and Ramseier, T. M., 1996. The catabolite repressor/activator (*Cra*) protein of enteric bacteria. *J. Bacteriol.*, **178**(12):3411.

[342] Gupta, S. and Clark, D. P., 1989. *Escherichia coli* derivatives lacking both alcohol dehydrogenase and phosphotransacetylase grow anaerobically by lactate fermentation. *J. Bacteriol.*, **171**(7):3650.

[343] Clark, D. P., Cunningham, P. R., Reams, S. G., Mat-Jan, F., Mohammedkhani, R., and Williams, C. R., 1988. Mutants of *Escherichia coli* defective in acid fermentation. *Appl. Biochem. Biotechnol.*, **17**(1-3):163.

[344] Mulquiney, P. and Kuchel, P., 1999. Model of 2, 3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: computer simulation and metabolic control analysis. *Biochem. J.*, **342**(3):597.

[345] Mulquiney, P. and Kuchel, P., 1999. Model of 2, 3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement. *Biochem. J.*, **342**(3):581.

[346] van Eunen, K., Bouwman, J., Daran-Lapujade, P., Postmus, J., Canelas, A. B., Mensonides, F. I., Orij, R., Tuzun, I., van den Brink, J., Smits, G. J., *et al.*, 2010. Measuring enzyme activities under standardized *in vivo*-like conditions for systems biology. *FEBS J.*, **277**(3):749.

[347] Cartwright, N., 1983. *How the laws of physics lie*. Cambridge Univ Press.

[348] Levins, R., 1966. The strategy of model building in population biology. *Am. Sci.*, **54**(4):421.

[349] Wiener, N., 1948. *Cybernetics; or control and communication in the animal and the machine*. John Wiley.

[350] Wolkenhauer, O., 2001. Systems biology: the reincarnation of systems theory applied in biology? *Brief. Bioinform.*, **2**(3):258.

[351] Ahn, A. C., Tewari, M., Poon, C.-S., and Phillips, R. S., 2006. The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Med.*, **3**(6):e208.

[352] Laughlin, R. B., 2006. *A different universe: Reinventing physics from the bottom down*. Basic Books.

[353] Rosen, R., 1958. A relational theory of biological systems. *B. Math. Biophys.*, **20**(3):245.

[354] Wingreen, N. and Botstein, D., 2006. Back to the future: Education for systems-level biologists. *Nat. Rev. Mol. Cell Biol.*, **7**(11):829.

[355] Pevzner, P., Shamir, R., *et al.*, 2009. Computing has changed biology—biology education must catch up. *Science*, **325**(5940):541.